DOCUMENT RESUME

ED 168 582 IR 007 068

AUTHOR Greenfield, Concetta C.; Serain, Daniel
TITLE Machine-Aided Translation: From Terminology Banks to Interactive Translation Systems.
INSTITUTION Carnegie-Mellon Univ., Pittsburgh, Pa. Dept. of Computer Science.; Institut de Recherche en Informatique et Automatique, Paris (France).
PUB DATE 77
NOTE 84p.; Appendix II not included due to illegibility

EDRS PRICE MF01/PC04 Plus Postage.
DESCRIPTORS Artificial Intelligence; Data Bases; Data Collection; Dictionaries; *Information Networks; *Language Classification; *Language Processing; *Machine Translation; Man Machine Systems; Programing Languages; *Programing Problems; Subject Index Terms; Telecommunication

ABSTRACT

The rapid growth of the need for technical translations in recent years has led specialists to utilize computer technology to improve the efficiency and quality of translation. The two approaches considered were automatic translation and terminology banks. Since the results of fully automatic translation were considered unsatisfactory by various reports, significant efforts were made to construct automatized terminology banks. A critical comparison of the main existing banks suggests that these banks are the first step toward interactive translation systems. Exchange problems in networks of banks and terminological cooperation between international organizations will necessitate standardization of terms, symbols, and definitions. It is proposed that automatized terminology banks be designed from a different viewpoint by making greater use of well known techniques in computer systems and artificial intelligence, based on evaluation of present and future developmental issues in data base organization, translation research, interactive translation systems, and man-computer communication. (Author/CWM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED168542

MACHINE-AIDED TRANSLATION: FROM TERMINOLOGY BANKS TO INTERACTIVE TRANSLATION SYSTEMS

Dr. Concetta C. Greenfield
Carnegie-Mellon Univ.

Dr. Daniel Serain
IRIA, Paris

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Douglas C. Coutts

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

IR00 7068

MACHINE-AIDED TRANSLATION :

FROM TERMINOLOGY BANKS TO INTERACTIVE TRANSLATION SYSTEMS.

C.C. GREENFIELD

Director of the Translation Center

Associate Professor

Carnegie-Mellon University

Pittsburgh, U.S.A.

D. SERAIN

IRIA/LABORIA

78150 Rocquencourt

FRANCE

Résumé

L'accroissement important du volume de traductions techniques de ces dernières années a forcé les traducteurs à utiliser les techniques de l'informatique afin d'améliorer l'efficacité et la qualité des traductions. Deux approches ont été prises : la traduction purement automatique et les banques de terminologie. Les résultats de la traduction automatique ayant été jugés insatisfaisants par le rapport ALPAC, d'importants efforts furent faits pour construire des banques automatiques de terminologie. Les principales banques existantes sont présentées et comparées d'un point de vue critique. L'approche suggérée est que les banques sont le premier pas vers des systèmes intéractifs de traduction. Ceci conduit à les concevoir avec une optique différente, et à utiliser des techniques bien connues en informatique système et en intelligence artificielle.

Abstract

The rapid growth of the need for technical translations in the recent years has led the specialists to utilize computer technology to improve the efficiency and the quality of translation. Two approaches were taken:automatic translation and terminology banks. Since the results of fully automatic translation were considered unsatisfactory by the ALPAC report, significant efforts were made to construct automatized terminology banks. The main existing banks

are presented and critically compared. The suggested approach is that these banks are the first step toward interactive translation systems. It is proposed that they be design from a different viewpoint by making greater use of wellknown techniques in computer systems and artificial intelligence.

Notes

. Ce travail a été supporté par l'IRIA (Institut de Recherche en Informatique et Automatique) et par les départements de Computer Science & Translation Center de Carnegie-Mellon University (Pittsburgh, U.S.A.).

. Ce rapport est également publié, en langue française, par le HAUT COMITE DE LA LANGUE FRANCAISE, sous la direction de M. Bruno de Bessé, Président de l'AFTERM.

MACHINE-AIDED TRANSLATION: FROM TERMINOLOGY BANKS TO INTERACTIVE TRANSLATION SYSTEMS

Contents

Part — Page

I. INTRODUCTION 1

The Explosion of Technical Terminology
New Facilities Offered by Terminology Banks
Machines and Translation, 1955-1976
  Mechanical translation
  Artificial intelligence
  Machine-aided translation

II. DEFINITION OF TERMS 5

Relevant Terms
The Study of Terminology

III. TERMINOLOGY BANKS, 1965-1976 9

General Characteristics
Data Acquisition and Storage
Access to the Data Base

IV. ESSENTIAL COMPONENTS OF A TERMINOLOGY BANK 15

Structure of a Record

Components of a Record
  Term specification information
    Entries
    Inverted entries
  Sense specification information
    Senses and meaning
    Definitions or contexts
    Matching contexts (or definitions) by translation
    Field classification

Organization of the Data Base of the CCE System, EURODICAUTOM
  Comments on the EURODICAUTOM data base
    Computer characteristics

Query Language

Making of Dictionaries
  Selection
  Inversions
  Synonyms
  Terminological digests

Part Page

Other Procedures
Conversion updating
Truncation function
Searching function
EURODICAUTOM classification function

V. EXCHANGE PROBLEMS IN NETWORKS OF BANKS 45

Networks of Banks
Data Structure Problems
Linguistic Problems
Redundancy Problems
Computer Networks

VI. FUTURE DEVELOPMENTS 48

Issues in Data Base Organization
The concept model
Indirect connections between terms
Precision in the request

Issues in Translation Research
Recommendations of the ALPAC report
Translation research at Carnegie-Mellon University

Issues in Interactive Translation Systems
Tanke: Artificial intelligence and machine aids to translation
Lippmann: Experimental on-line computer aids for the human translator
Kay: An incremental approach to aids to translation
The TARGET prototype system

Issues in Man-Computer Communication
A testable system
A flexible approach to system design
Interactive facilities

VII. APPENDIX I 70

VIII. APPENDIX II. 71

IX. REFERENCES 74

I. INTRODUCTION

As the technological explosion continues and interdependence among nations grows, the proliferation of terms and the cost of high-quality translation have led to the establishment of a number of terminology banks. Every time the idea of a terminology bank is broached, the question arises: why not a dictionary? To answer this question we need to refer to two sets of data: the first set describes the explosion of technical terminology in the last ten years; the second set contains information about the unique facilities offered by a terminology bank.

The Explosion of Technical Terminology

In the past 15 years, activity in research and technology has increased phenomenally. Scientific publishing has increased proportionately, and will continue to increase. By 1987, the annual production of technical and scientific information is projected to be at least six times the current rate[cf., And, p. 9].

To keep track of new terms, it has become commonplace for companies to collect glossaries of their own terminology. [cf., Spe] This is particularly true for the vocabulary of oil, which has been compiled largely by the efforts of oil companies. The Shell Oil Company of Egypt, for example, has produced glossaries of technical terms, and so has the Arabian-American Oil Company.

New technical terms are coined daily to deal with developments in research and technology. How are these terms handled for translation purposes? So far, the only widely available method is by inclusion in a printed dictionary. This is unfortunate, because the interval between dictionary publications is long. A study of the publication of French-English dictionaries, done by Carnegie-Mellon University, reveals that even in a fast-moving field such as electronics, bilingual dictionaries in French and English appear only on the average of every 2.4 years. [cf., Voj] By the time most terms get into a dictionary, they are no longer the subject of research - the period when they were discovered, developed and publicized is already over. The time translators need them most is past. On the other hand, glossaries collected by private companies are used only for internal purposes. Since they are intended primarily for their compilers, no comparison is made with terms of other organizations; often different organizations use different terms for the same process.

Unlisted new technical terms impede the process of translation each time they are encountered. In order to translate an unlisted term, a translator must search the literature himself, ask his friends, or make an informed guess. This leads to inaccuracies and duplications of effort. "It has been claimed that up to 60% of a conscientious translator's work time is consumed in tracking down proper terminological information." [cf., Lip]

It is obvious that a need exists for the documentation and control of the enormous quantity of terminology produced recently. However, there is also a need for some kind of regulation of quality. A number of terms which take on different shades of meaning in different fields may tend to confound the translator. For example, he may find the equivalent of a target term in a legal dictionary which does not fit his particular field or context, but because of the time

and expense of the additional search, he may be tempted to make do with an approximation to the correct equivalent. The Bundessprachenamt found that the amount of mistranslated terminology increases proportionally to the level of technical difficulty of the source text. Terminological inaccuracies account for 62.1% of the total translation errors (where the range of error types includes such categories as orthographical mistakes, punctuation, wrong prepositions, inflectional errors, foreign word order, text inaccuracies, and omission of information). [cf., Kre2]

It was precisely in response to the growth of technical terminology that the idea of computerized terminology banks has gained increasing popularity across the world, generating considerable amounts of experimentation and research. It has became clear that by using these computerized systems, large amounts of terminology can be collected, stored, and made available more rapidly than through traditional methods of dictionary publication, in which a finished corpus is published every 2.4 years.

New Facilities Offered by Terminology Banks

It soon became apparent that terminology banks offer facilities which dictionaries cannot offer. For example, computer programs may not only be used to store, sort, and interfile new terms; they may also be used to scan a journal article, pick up selected technical terms, and produce a list of these terms in the sequence in which they occur, along with their target language equivalent. Erhard Lippmann calls these lists "terminological digests." [cf., Lip] In an experiment, translators using this type of list took 1/3 less time and made 2/3 the errors of translators using conventional translation aids, such as printed technical dictionaries. [cf., Nat]

Another facility available in a terminology bank is the instant composition and production of specialized glossaries of terms related to a field or subfield. This facility is useful for clients who want to order a glossary in specialized fields. A terminology bank is like a dictionary whose pages can decrease or stretch, contract or expand, according to the immediate needs of its users.

In fact, the facilities offered by existing terminology banks are adequate to control both the quantity and quality of the avalanche of new technical terms. However, further improvements may be added to facilitate other aspects of the process of translation. These improvements include procedures for:

1. Retrieving contexts of all occurrences of a term in the source document simultaneously. This eliminates the repetitive lookups which occur each time the term is encountered in text. It also eliminates backtracking to amend early equivalents when better ones are discovered later on in the translation.

2. Printing out the edited target text. This eliminates the final typing of the finished translation, a particularly acute problem since translators are seldom provided with bilingual secretaries and must spend much time checking and correcting what is typed for them.

b. These procedures may be incorporated into a terminology bank by implementing them in a translation system which allows for interaction with a dictionary and editor. The translator would work on-line, editing the source and target text, assisted by a potentially enormous multilingual dictionary accessible far faster than the large collection of hard-copy dictionaries possessed by most translators.

Machines and Translation, 1955-1975

Research in the last ten years indicates that an interactive translation system of this type is less ambitious but certainly more realizable than the fully automated translation which was attempted in the early 1960's. In fact machine-aided translation is an outgrowth of the research and implementation of that period. A brief survey is presented below.

Mechanical translation

From 1955 to 1965, there was an extensive attempt to develop mechanized translation procedures. These procedures operated entirely without human intervention from the time the original source text was read into the computer until the final translation was printed out. However a representative output of a mechanical translation procedure indicates the result to be an incomplete translation requiring extensive post-editing. (Machine translated samples quoted in the report are presented in Appendix I.) Part of this lack of success was due to ignorance about the process of translation. Mechanical translation was laid to rest by the report of a committee of the National Academy of Sciences in 1965. [cf., Nat] The report stated that results in the field of mechanical translation were not encouraging in terms of cost effectiveness due to the extensive post-editing. However, the report encouraged the use of computers as aids to human translators.

After this report appeared in 1965, the research energy invested solely in mechanical translation was rechanneled into three separate tasks: (1) the improvement of existing completely mechanized procedures (mechanical translation), (2) the development of small, completely mechanized simulations of language processing (artificial intelligence), and (3) the development of aids to translation (machine aided translation) via terminology banks.

Mechanical translation systems after 1965 produced output very similar to that of the earlier systems. J. Schultz from the Siemens Terminology Bank concluded in a 1975 paper [cf., Sch]:

> Previous efforts to carry through translations by means of computers have not led to results which permit replacing man by computers, or even allow us to anticipate this in the foreseeable future. However, the continually increasing flow of scientific, technical and commercial information, which is to be continually translated, makes it appear necessary and meaningful enough to employ the computer as an aid for the translator.

According to Chuck Bieger, commenting at the Foreign Broadcast Information Service Seminar in March 1976 [cf., Bie]:

> Almost no theoretical advances have been made since the late 50's and early 60's - at least as I have understood the state of MT was at that time. Specifically, all the systems still rely on basically a word to word translation, with some syntax gratuitously tossed in, and dressed up with so-called "micro-glossaries." No system apparently copes with (1) reference, (2) even the most rudimentary model of what it is that's being translated, (3) contextual usage of words (or even automatic micro-glossary sifting), (4) semantics...or (5) "understanding." Basically, the quality of the dictionary is how cleverly it has been tuned to a par-

ticular application. While this might produce results adequate for a very technical translation to audiences who already know most of what is being said, it falls about ten badly-needed theoretical advances short of so-called "quality" MT, or even low quality MT in less technical areas such as literature or political science. In short, there is every reason to be disappointed with the showing.

Proponents of these systems, however, say that their users are satisfied with the post-editing output. Ninety-two percent of the persons responding to a survey at Oak Ridge National Laboratory judged their machine translations to be "good" or "acceptable." At the same seminar, Paul Chapin stated [cf., Cha]:

> Researchers who start from a scientific orientation...will not or cannot accept the notion that mechanical translation has succeeded. They are defining a system "works" as a system "equals or betters human performance," which is the scientifically interesting sense of the term, but clearly too stringent a definition.

Artificial Intelligence

While those working with mechanical translation systems have attempted to improve performance by perfecting existing systems, other researchers have sought to produce eventual improvements in automatic translation by trying to simulate the process of language comprehension itself. To achieve this goal, some of these researchers are trying to design a representation for deeper conceptual structures which can be used in the process of making inferences. [cf., Wil] Other researchers are trying to construct small systems which perform complex operations of intelligence in a restricted environment. A famous system of this type is Winograd's "block world" in which a computer converses with a human in good English, about an arrangement of toy blocks. [cf., Wil] The computer is able to remember and infer facts about its universe, the arrangement of the blocks. Hopefully, what makes such a small system perform will also make a larger system perform. Critics of this approach doubt that this is possible. How can the masses of cultural information necessary for such a large, functioning system every be isolated, analyzed, and stored?

Machine-aided translation

Supported by the ALPAC report recommendations, energies have been shifted to produce even more refined computer aids to help the translator. "Between 1965 and 1968, there arose the concept of a financially self-supporting, computerized word-bank, based on independent lexical research and designed to furnish the translator with specialized vocabulary couched in a specific context." [cf., Par2] Outside the U.S.A. at least 16 terminology banks offer batch or on-line automated dictionaries to a variety of users. In the last years considerable interest to establish a terminology bank in the U.S.A. has been expressed by several organizations and researchers. In particular, S. Lipmann, Chairman of the Committee for Automatic Translation of the ATA (Association of American Translators) and M. Kay (RAND, Palo Alto) [cf., Kay] have noticeably advanced the state of the field by suggesting interactive systems that offer text processing facilities in addition to on-line access to the bank [cf., Part VI.] These recommendations are taken into account in the design and implementation of an interactive system called TARGET at Carnegie-Mellon University [cf., Crem, Cre].

II. DEFINITION OF TERMS

Machine-aided translation is an interdisciplinary field. It spans the disciplines of translation, lexicography, and computer science, and related them to the field of terminology. It is thus important at the outset to define the sense in which some technical terms will be used in this paper, and to explain major issues in terminology.

Relevant Terms

Context is the part of the text characterizing a term which surrounds the term and whose meaning and value depend on it.

Entry is a word, a compound word or a group of words which are used to access a record.

Fiche is the printed form on which a term and its accompanying information is written, before entry into the data base.

Field is a term designating a more or less extensive field of knowledge or scientific or technical activity.

Field of a record is the access unit of a computerized data base. A record is a set of fields.

Keyword is a one-word entry included in a multi-word term.

Lexicography is the science of compiling glossaries, lexicons, and dictionaries. (F... Zgusti and Rey)

Lexicology is that part of linguistics that is devoted to the study of lexical facts.

Record is the term and its accompanying information. This was originally written on a fiche.

Term bank is the delineation of the meaning of a word by either its direct and referential equivalency between two or more languages.

Term is the word or phrase for which the record is written. It is an entry.

Thesaurus is a dictionary of synonyms.

The Study of Terminology

It should be noted that the field of terminology has its own problems distinct from that of translation. Terminological research is usually important to subject specialists, translators, interpreters, information specialists, and all those involved in technical and professional communication. Issues studied in terminology may be specifically applied to translation. The International Information Centre for Terminology (Infoterm) has defined ten meanings for terminology, namely:

Terminology: a science. This is the field of knowledge covering the formation and naming of concepts, either in a special field or in the aggregate of all subject fields [cf. ... text]. "The theory of terminology is an ideal branch of science dealing with the laws of specialized languages while linguistics deals only with the laws of common language." [cf. ... Rey]. The study of terminology is a transdisciplinary field developed from linguistics, logic, ontology, information sciences and the specialized subject fields.

Terminology: the aggregate of terms representing a system of concepts particularly in a special field. [cf. ... text]

Robert Dubuc of the Former Montreal Bank offers a working distinction between the meaning of terminology and the term terminology as they are used in the environment of terminology banks:

> Le terminographe s'emploie à définir les termes, la terminologie s'occupe à les découvrir en situation concrète de vie, à les promouvoir, selon le cas, en interrogeant, si l'on peut dire, une banque de l'usage que vont les divers locuteurs: écrits, documents, manuels, notices, etc, ou bancs, les utilisateurs eux-mêmes. [Le particulier "le terminologie consacre" [cf. ... Dub ..., p. 1 (would be terminology "une plus criteria were le terminographe sait vers les besoins spécifiques des traducteurs" [cf. ... Dubc, p. 3]

Terminological research is being carried out internationally. Centers of interest study of terminological problems are the Canadian universities and the Bank de la Langue Française in Quebec. Fundamental work in theory of terminology is also done at the Academy of Sciences and at different universities in the USSR (Leningrad and Moscow). Recently Bruno de Bessé, (head de stratégie experte du Haut Comité de la Langue Française and President of the Association Française de Terminologie, organized the Ultième International de Terminologie (International Congress of Terminology) in Paris (1978) which surveyed existing terminological problems.

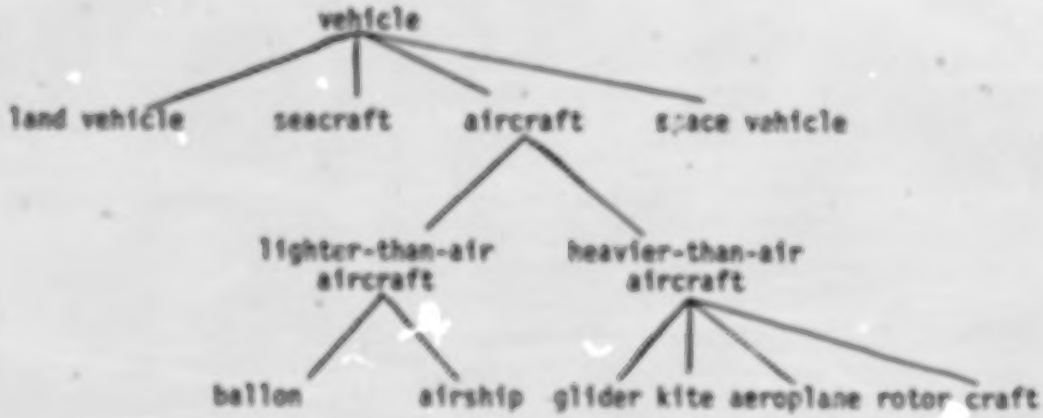
Finally Robert Dubuc of Infoterm has published an outline of a study on the development of a network of terminology-processing banks and terminology documentation centers, documenting terminological principles developed on the basis of (3) recommendations. Such terminological and terminographical principles were applied in their entirety for the first time in the Eltic project "The Machine Test" undertaken by Professor Bachar on behalf of the United Nations Economic Commission [cf. ... Bach]. Before his death, Professor Bachar was an outstanding author in the general theory of terminology.

These two ideas developed in the field of terminology directly affect the organization of computerized terminology banks: (1) the distinction between a term and a word in everyday language and (2) concept relations for the construction of concept classifications.

The idea of the distinction between a term and a word in everyday language is discussed by Olga Akhmanova and Galina Agapova. [cf., Akh.] They establish the basic criterion of "definability." The words "barn," "cloud," and "fun" are part of the ordinary language: the fact that they belong to a sphere of human experience does not make them terms. They are considered terms only when an exact scientific definition is attached to them, in which case they become part of a different "sub-system" of the vocabulary. For example, if "soap" is defined as a substance for washing, it is an everyday word because the parameters of its definition are from everyday vocabulary; if it is defined as a salt of a fatty acid, it is a term, because the parameters of its definition are from chemistry.

The idea of concept relations is discussed by Helmut Felber. [cf., Fel5] He stresses that in the theory of terminology, the value of concepts is independent of the realm of terms. There may be one term (ex. pipe) to which correspond three concepts. Terminology work must start "from concepts and not from terms." Sometimes terminology work is started from a collection of terms; these terms are defined and later arranged in alphabetical order. However, very soon it is found that some definitions overlap or contradict each other. There is more than one type of concept relation. We can have (1) logic relations (genus - species • fruit ≡ apple; coordination • apple ≡ pear), (2) ontological relations (whole-and-part • eye ≡ eyeball; coordination • eyeball ≡ eyebrow), and (3) connecting relations (determination • vehicle • motor • motor vehicle). The following is an example of a genus-species system. [cf., Fel5, p. 9.]

Figure 1. Example of a Genus-Species Concept Relation



Concept relations are used particularly for thesauri. There has been little discussion as to whether it would be more efficient to organize a word data base according to a system of concepts rather than as a dictionary. The strongest point in favor of the concept system is that the information in most entries (the equivalent, the context or the definition) does not provide the translator with other related words in the same concept system. It is assumed that the translator already knows related words which may be useful references, but very often he does not. Yet the efficiency of the bank would be greatly increased if he had access to all other terms related by the concept system involved. Only the ArNOR bank in Paris uses this system.

III. TERMINOLOGY BANKS, 1965-1976

General Characteristics

Translation aid systems can be bilingual or multilingual, accessible on-line, or by batch processing. The chart in Figure 1 offers an exhaustive presentation of existing banks, outlining fields and types of terms, types of service, languages, entry information, subject classification, and computer programs.

A few useful inferences can be made from this chart:

1. Usually, stored terminology is technical, not general.
2. Users are terminologists, translators, and linguists.
3. Predominant fields are science and technology, closely followed by economics.
4. Common languages are English, French, and German, followed by Russian.
5. Definitions are used more often than contexts to characterize terms.
6. A unique field classification is adopted by each bank.
7. The main use of the banks is off-line with limited on-line access.

All banks are based on similar assumptions which can be listed in order of decreasing importance:

1. To improve the quality of translation by supplying the user with the most updated, current, pertinent answer to his particular terminological question.

2. To save time and hence boost productivity by making a correct expression available to all users in a minimal amount of time.

The use of the bank determines its organization, design, and implementation. For instance a bank designed for the purpose of bibliographic documentation is organized differently from one that serves the needs of translators. This is because bibliographical and linguistic references, while enriching the process of documentation, often prove cumbersome during translation. More specifically:

1. Extensive documentary information is generally not useful to the translator.

2. While exhaustiveness is the major criterion for documentation, relevance is the major criteria for translation.

3. A documentation system must produce a large amount of information for one entry; a translation system must produce relations between many entries.

4. A translation system must be able to respond immediately, while a translation is in progress; this speed is not as necessary for a bibliographic documentation system.

Data Acquisition and Storage

Generally, data acquisition techniques are dependent on term sources. A terminology bank uses a variety of term sources:

1. Standardized terms available from standard technical dictionaries.

2. New terms used on site, obtainable from current translation work and the private files of translators (see Part IV); new terms from contemporary journals and technical reports.

3. Neologisms coined by terminologists.

The need for finding new terms which are not yet in a dictionary but are used on site during the daily transactions of many organizations must be stressed. The collection of this "underground" terminology should be a primary feature of a bank; for instance, the usage of the word "runner" originating in a Youngstown (USA) steel mill does not appear under the corresponding German entry in current steelmaking dictionaries, yet this usage has spread to quite a few steel plants in other areas of the USA. These are the kind of terms every bank should search for.

After data is acquired it is stored in the bank. There are three common methods of adding terms to the data base:

1. Manual entry of terms (through terminals and punched data cards).

2. Automatic generation of terms

   a. If terms are available on magnetic tapes (from book distributors or other terminology banks) they can be exchanged and then converted to the particular format used by the bank. Such an exchange operation has been successfully accomplished between the Terminology Bank of Canada and the CCE Bank.

   b. Terms which are typed on paper with a special character set can be scanned by an optical reader (as at the CCE Bank).

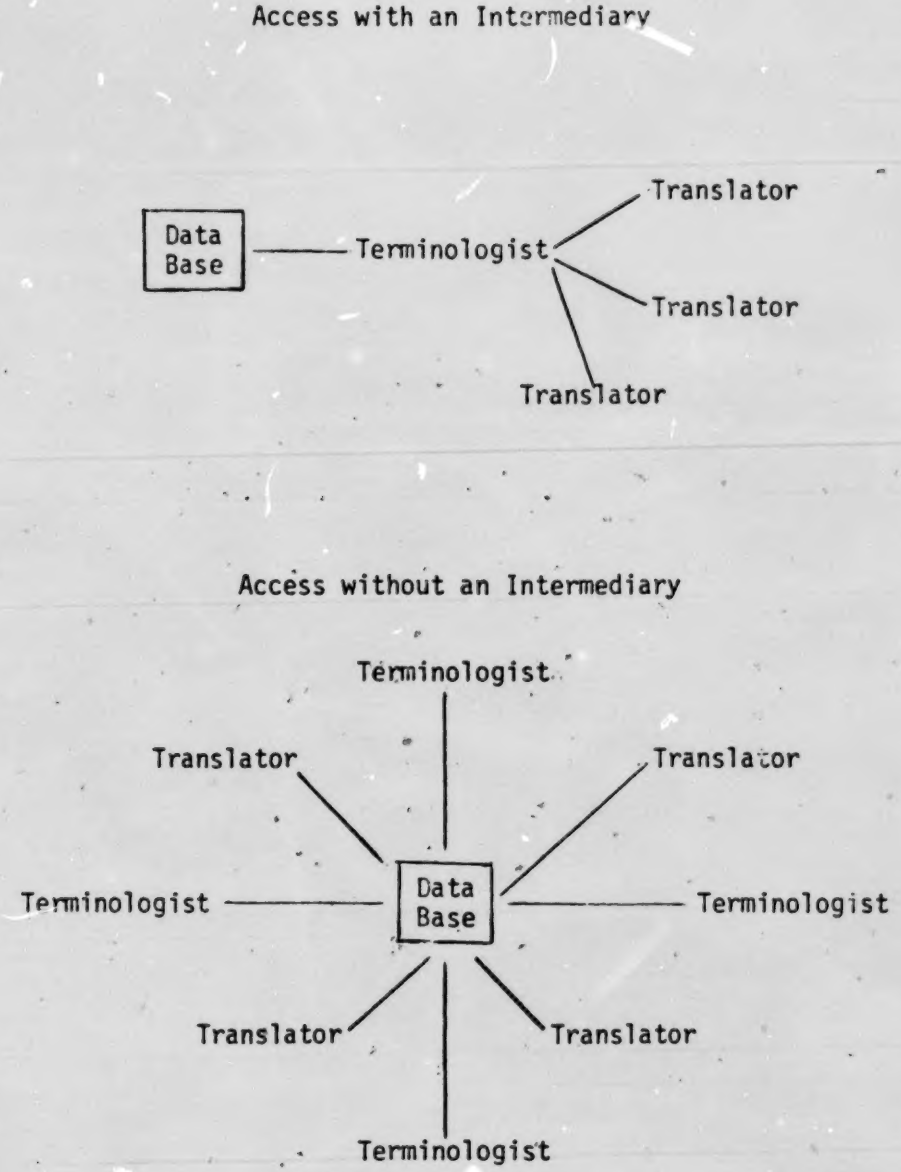
Generally the acquisition of terms is manual while the input is automatic. However, if a text is stored in the memory of the computer, the selection and acquisition of the new terms can be also automatic (see Part IV).

Access to the Data Base

Access can be on-line (direct) or off-line (indirect). Direct, on-line access is the most efficient method of retrieval because it allows interaction between the user and the bank. This method lets the user take advantage of other facilities; for example, the procedures for creating and editing terms, scrolling through the dictionary, or for creating his own intermediate file.

While some banks have facilities for direct access, none of them make these facilities available directly to translators. Instead, specialized terminologists act as intermediaries between the data bank and the users. This is illustrated in Figure 2. The terminologists access lists of terms needed by the translators; these terms are then returned to the translators who incorporate them in their target text. Two reasons are usually presented to justify the lack of direct access by users: one is cost, the other is that users (especially translators) dislike computers. While it is clear that computer costs remain more or less the same whether a terminologist or translator is accessing the terms, resistance to computer technology can be decreased if systems are designed with the user's attitude in mind (see Part VI).

Figure 2. Access With and Without an Intermediary



IV. ESSENTIAL COMPONENTS OF A TERMINOLOGY BANK

Structure of a Record

A terminology bank is composed of a set of records and of mechanisms to access, create, and update them. "Fiche" is the format for the entry of a term into the data base. The information on a fiche constitutes a "record". The record is the logical unit containing data needed by the users of the bank.

The structure of a fiche is determined by the objectives of the data base. Figures 1, 2, and 3 present three interesting fiches from the Montreal, the Quebec, and the CCE Banks.

Figure 1. Fiche of the Montreal Bank
(Now part of the Terminology Bank of Canada)

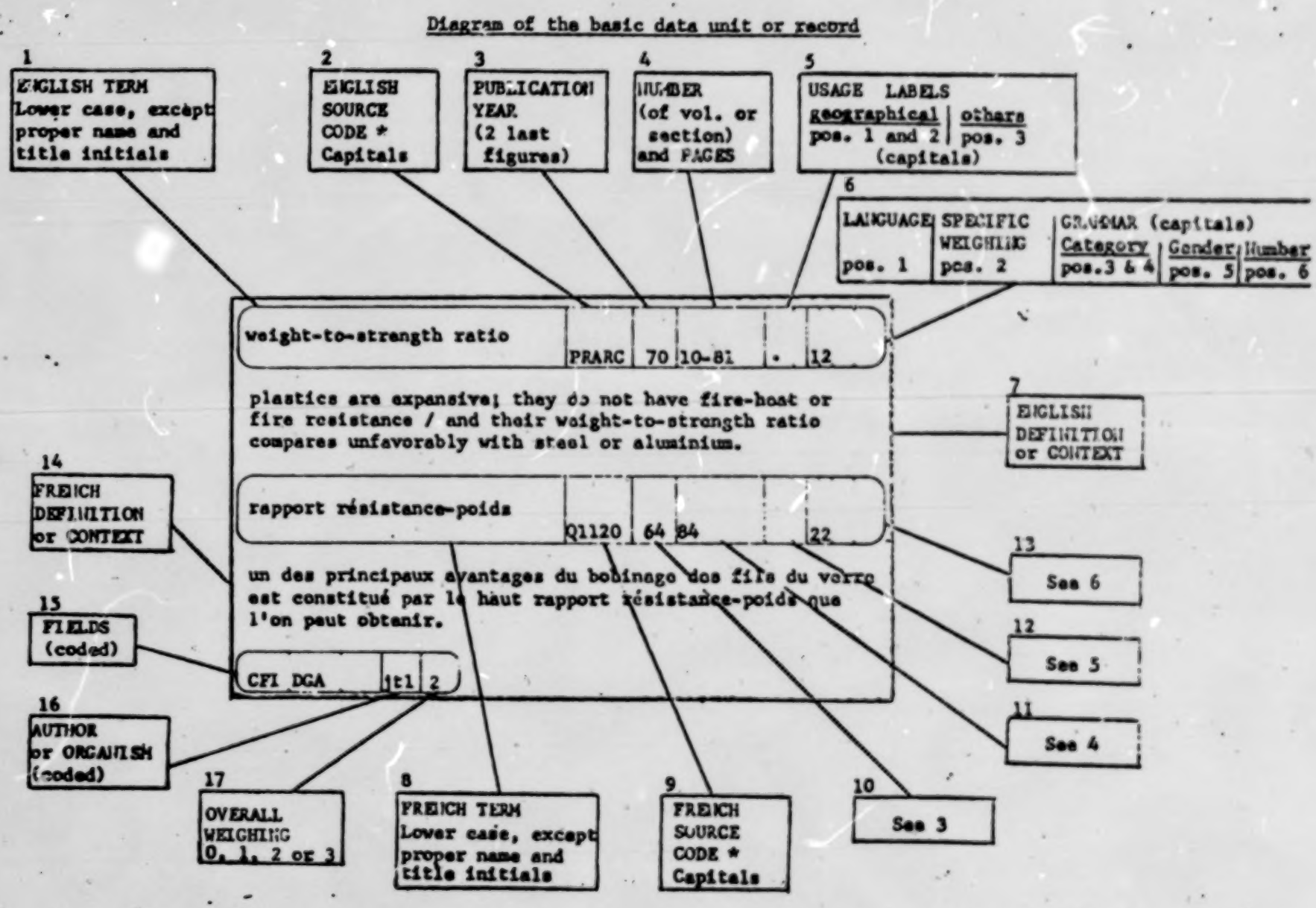
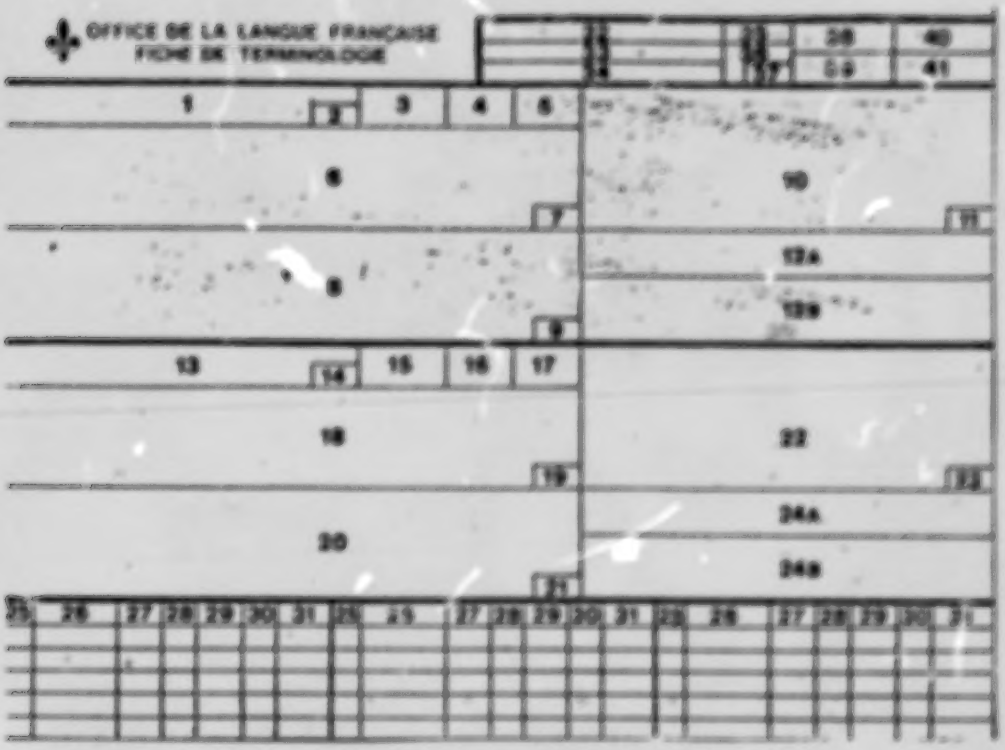


Diagram of the basic data unit or record
1
ENGLISH TERM
Lower case, except proper name and title initials
2
ENGLISH SOURCE CODE * Capitals
3
PUBLICATION YEAR (2 last figures)
4
NUMBER (of vol. or section) and PAGES
5
USAGE LABELS
geographical pos. 1 and 2 | others pos. 3 (capitals)
6
LANGUAGE pos. 1 | SPECIFIC WEIGHING pos. 2 | GRAMMAR (capitals) Category pos. 3 & 4 | Gender pos. 5 | Number pos. 6
weight-to-strength ratio
PRARC 70 10-81 . 12
plastics are expansive; they do not have fire-heat or fire resistance / and their weight-to-strength ratio compares unfavorably with steel or aluminium.
7
ENGLISH DEFINITION or CONTEXT
14
FRENCH DEFINITION or CONTEXT
rapport résistance-poids
Q1120 64 84 22
un des principaux avantages du bobinage des fils du verre est constitué par le haut rapport résistance-poids que l'on peut obtenir.
13
See 6
15
FIELDS (coded)
12
See 5
CFI DGA 1t1 2
16
AUTHOR or ORGANISM (coded)
11
See 4
17
OVERALL WEIGHING 0, 1, 2 or 3
8
FRENCH TERM
Lower case, except proper name and title initials
9
FRENCH SOURCE CODE * Capitals
10
See 3

* In areas 2 and 9, the sign "@" is the code symbol for the Banque de terminologie

720927

Figure 2. Fiche of the Banque de Terminologie de Quebec



1 ENTRÉE DANS LA LANGUE DE DÉPART 13 ÉQUIVALENT DANS LA LANGUE D'ARRIVÉE
14 19 21 23 APPEL(S) DE RENVOI À LA(AUX) SOURCE(S)
15 INDICATIF(S) DE GRAMMAIRE 16 INDICATIF(S) DE LANGUE ET DE PAYS
17 COTE DE PONDÉRATION 18 DÉFINITION 20 CONTEXTE 22 NOTE
24 RENVOIS A) À DES FICHES SECONDAIRES: VARIANTES ORTHOGRAPHIQUES (VO), ABRÉVIATIONS (ABREV), SYNONYMES (SYN), TERMES APPARTENANT À UNE AUTRE LANGUE QUE LA LANGUE DE DÉPART OU D'ARRIVÉE (AL), DÉFINITION (DEF), CONTEXTES (CONT), NOTES (NOTES)
24 RENVOIS B) À DES FICHES PRINCIPALES: RENVOIS "VOIR" (V), RENVOIS "VOIR AUSSI" (VA)
25 INDICATIF DU RENVOI DE LA SOURCE 26 N° DE LA SOURCE
27 ANNÉE 28 VOLUME 29 NUMÉRO
30 MOIS OU SAISON(S) POUR UN CATALOGUE 31 PAGE
32 DOMAINE D'EMPLOI 33 1er SOUS-DOMAINE 34 2e SOUS-DOMAINE
35 INDICATIF DE L'ILLUSTRATION: ASTÉRISQUE
36 N° SÉQUENTIEL DÉFINITIF DE L'ILLUSTRATION (POUR FIN DE PUBLICATION)
37 CODE D'IDENTIFICATION DE L'AUTEUR DE LA FICHE ET DE L'ORGANISME AUQUEL IL APPARTIENT, S'IL Y A LIEU
38 MOIS ET ANNÉE (LES DEUX DERNIERS CHIFFRES DU MILLÉSIME) DE LA RÉDACTION DE LA FICHE ET DES MISES À JOUR
40 N° SÉQUENTIEL DE CLASSEMENT PROVISOIRE DE LA FICHE
41 N° SÉQUENTIEL DE CLASSEMENT DÉFINITIF DE LA FICHE

Figure 3. Fiche of the CCE Bank
(Some information has been entered)

$ NI (NX)			$ TY	VIN 63	$ DE	BTL	
$ CF		$ AN		$ AU			
$ CM							
$ DG	$ PN	$ DF					

vides existant dans une plantation par suite de la non reprise de boutures ou d'accidents

$ VE	$ MC	

nanquants

$ NT

$ RF	office international de la vigne et du vin

$ FG	$PN	$ DF

vuoti che si verificano nei vigneti di nuovo impianto in seguito alla morte delle giovani piantine

$ VE	$ MC

fallanze

$ NT

$ RF	office international de la vigne et du vin

$ EG	$ PN	$ DF

im Weinberg diejenigen Stellen, in denen Pflanzreben nicht gewachsen oder durch irgendwelche Umstände zugrunde gegangen sind

$ VE	$ MC

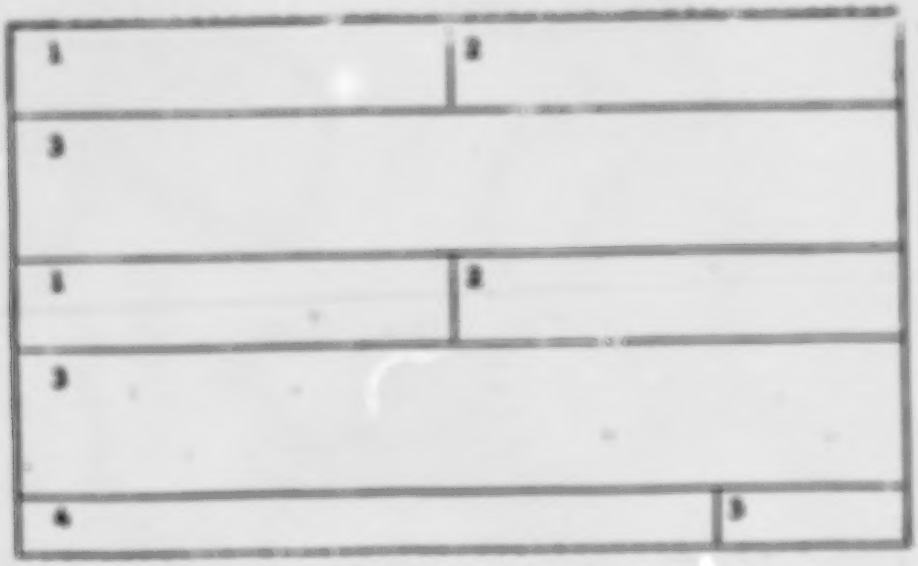
fehlstellen

$ NT

$ RF	office international de la vigne et du vin

Irrespective of different formats, there are a number of items that all fiches have in common. The FIT (Federation Internationale des Traducteurs) in May 1978 recommended a "universal format." This fiche is quite streamlined, containing only information assumed to be directly relevant for translation. It is illustrated in Figure 4.

Figure 4. The "Universal Format"



75 mm x 125 mm

1 Semantic unit
2 Source(s)
3 Definition(s) or significant context(s)
4 Field(s) of application
5 Author of information unit

In the following section we will discuss these and other items of information which may be included in a fiche. They are the components of a term's record, i.e., the information that describes the term.

Components of a Record

There are two sets of information in a record. One is related to term specification, which includes information pertaining to all meanings of the term; the other is related to sense specification, which includes information relevant to individual meanings. Some of the more important information categories of term and sense specification will be discussed.

Term specification information

This information includes:

- term (consisting of one or more entries)
- language
- identification number
- place of origin of the record.

Entries

An entry is defined as the entire term (word, compound word, or phrase) or any part of it which gives access to the term. Thus, the term "black off valve" has the entries "black off valve," "black" and "valve," since typing any of these three items accesses "black off valve." "Black" and "valve" are considered "key words" of "black off valve." (cf., Hall). The term is stored with the accompanying information in the internal data term; each of the information categories is made into a separate field. The fields can be accessed separately. The more access fields per record a terminology bank has, the more sophisticated it is, since this allows the user to interrogate it through any of these fields. In practice, however, the number of fields which may be accessed per record must be limited since it is costly to implement an access method for each field.

In the Montreal Bank, the entire term may be accessed. In the EEC Bank, not only are key words accessible, but the context of a term itself contains multiple entries. This means that the term can also be accessed through words, compound words or phrases of its context. Thus, when a term is typed in, the answer is composed of all records in which it appears as a main entry or as part of the context field (cf. Fra; cf. Hall). In the SPEER Bank, which is a documentation system (discussed below), each term is described by a set of "basic terms" belonging to a thesaurus. Use of these basic terms as supplemental entries allows the user to find the term. This feature is useful in a documentation system in which users have to find the appropriate term in a given language through its synonyms (cf., Lau).

Inverted entries

Entries may be reversed to produce inverted entries. This means that information contained in an entry made with English as the source language and French as the target entries equally well in French as the source language and English as the target. Reversibility is a never-to-be-attained ideal in the field of translation. However, some information is usually reversible. It can be inverted automatically from the original entry and supplemented by a terminologist. Figure 3 lists the information which can be legitimately inverted to produce machine-generated entries.

Figure 6. Categories for Entry Inversion

Original Entry (Correctly/not Generated)

term	capital letter
language of term	EN
source of term	F
grammatical category	noun
Field classification	A.01
equivalent	letter molecule
language of equivalent	EN
source of equivalent	A
date	2

Inverted Entry (Machine Generated)

term	letter molecule
language	EN
source of term	A
grammatical category	noun npr
Field classification	A.01
equivalent	capital letter
language of equivalent	EN
source of equivalent	F
date	2

Sense specification information

This information includes:

- equivalent of the term
- definition of the term and its source
- context of the term and its source
- subject field of the term
- generic terms
- specific terms
- grammatical categories (parts of speech, gender, etc.)
- abbreviations, orthographic variations
- reliability code for term
- name of author
- year of creation
- comments

The information under a sense of a term delimits its meaning for an exact and reversible equivalency between it and a term sense in another language. The process of delimitation is sometimes negative, when overlapping meanings of a word are stripped away and considered singly; sometimes it is positive, when additional information is supplied to make up for semantic deficiencies.

Figure 6 contains a number of terms found by tracing equivalents of the English term "blade." The lines on it represent equivalences between the term and its equivalents. They are referred to as "links" or "equivalence-links" in what follows. In preparing Figure 6, we followed every term through its equivalents until a dead end was reached.

It will be noticed that the German word "Schaufel" and the French word "aube" are equivalents of both "blade" and "vane"; however, this is not true of any of the other equivalents of either "blade" or "vane." In order to account for this, we posit that both terms have more than one sense; there are as many senses of a term as are needed to provide the correct equivalence classes for translation purposes. "Blade" and "vane" need at least two each, as may be seen by inspection of Figure 6.

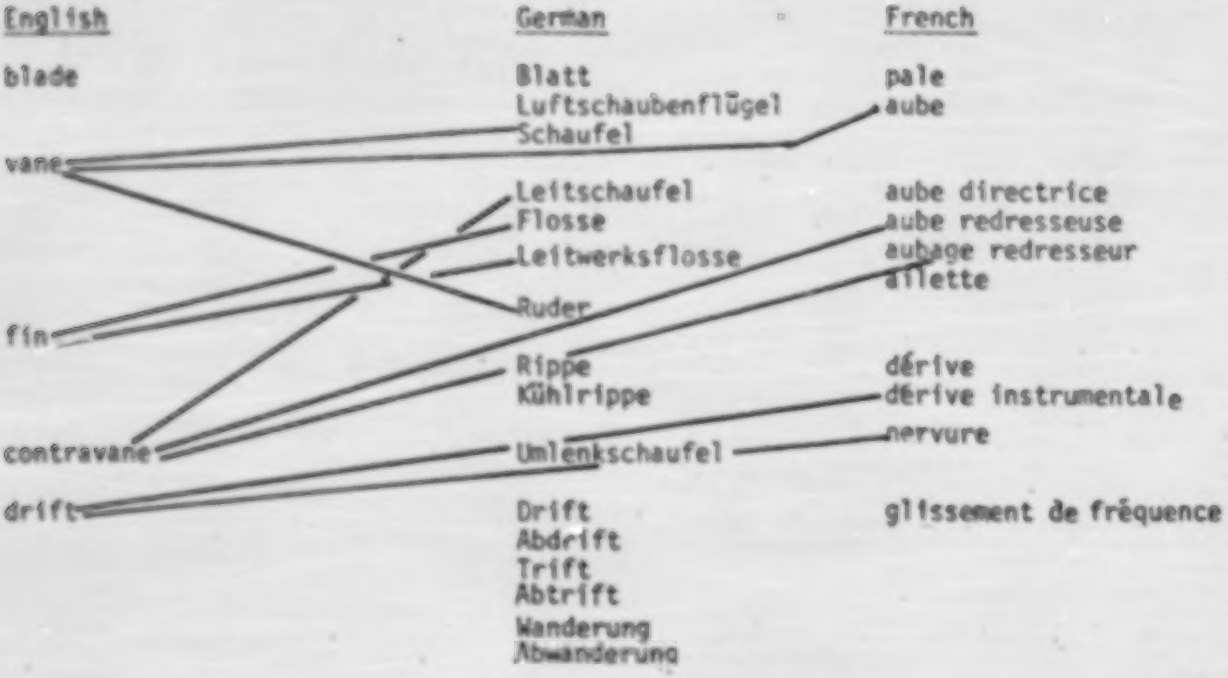
How can these senses be distinguished? In some cases, a general semantic code such as the "field of classification" may suffice, but in the case cited above, it does not. Another possibility is the use of finer semantic distinctions like definitions for the purpose of classifying words in the data base and for the purpose of choosing correct equivalents.

Senses and meaning

Terms usually have more than one meaning. This is indicated by multiple definitions. How should these meanings be handled in a terminology bank designed for translation?

Figure 6. Linking Equivalents of "Blade"

All Equivalents, and Equivalents of Equivalents, of "Blade"
(Links between Equivalents are Indicated by Lines)



From International Academy of Astronautics. Astronautical Multilingual Dictionary of the International Academy of Astronautics. (New York: Elsevier, 1970).

Multiple meanings of one term may vary on two parameters:

1. the multiple meanings take the same or different equivalents in a target language;

2. the multiple meanings refer to similar or quite different entities.

a. The word "solution" has two meanings defined by the two following definitions:

   i. dissolution, i.e., the mixing of a solid liquid or gaseous substance with a liquid, forming a homogeneous mixture from which the dissolved substance can be recovered by physical processes.

   ii. the homogeneous mixture formed by the operation of solution.

      1. The multiple meanings take the same equivalents in the target language, French (both "solution" and "dissolution").

      2. The multiple meanings refer to similar entities.

b. The word "table" has two meanings defined by the two following definitions:

   i. a piece of furniture which is a smooth flat slab supported by four legs

   ii. a systematic arrangement of data.

      1. The multiple meanings take the same equivalent in the target language, French ("table").

      2. The multiple meanings refer to quite different entities.

c. The word "bolt" has two meanings defined by the two following definitions, and illustrated by a list of relevant terms in which "bolt" is a key word:

i.	manual lock	foot bolt	
		head bolt	
		lever operated bolt	
		chain pull operated bolt	
		sliding bolt	
		spring bolt	
ii.	fastener	rib bolt	machine bolt
		turned bolt	roofing bolt
		unfinished bolt	grange bolt
		hook bolt	anchor bolt
		hex head bolt	toggle bolt

1. the multiple meanings take different equivalents in the target language, French (manual lock = "verrou"; fastener = "boulon", "cheville")

2. the multiple meanings refer to quite different entities.

One way to handle this situation is to group meanings into "senses." Meanings in each "sense" take the same equivalents. Thus, "solution" and "table" would each have only one sense, while "bolt" would have two. Although this plays down the distinctions between entities, which will be described only by multiple definitions within a sense, grouping meanings by equivalents is convenient for translation purposes.

Definitions or contexts

Although definitions are important conveyors of the meaning of a term, there is no general agreement as to whether they convey it with maximum efficiency. This is part of a broader issue that has been encountered in the field of lexicography. Here, term meaning is conveyed in multilingual dictionaries as well as terminology banks by different combinations of (1) equivalent terms, (2) definitions and (3) contexts. What is undetermined is the relative efficiency of each of these devices in conveying the term's meaning. The conclusion of a survey investigating the practice of existing multilingual dictionaries is:

> We can conclude that lexicographers agree that equivalent terms are necessary to convey term meaning but disagree on the value and relative importance of definitions or usage samples.
>
> In the case of terminology banks, the term entry formats also consistently use equivalent terms but are divided on the use of definitions or usage samples. Information for each term is presented monolingually, along with a list of locations where equivalent term entries in other languages are stored within the computer. [cf., Mor]

Figure 7 shows the use of definitions and usage samples by four terminology banks.

Figure 7. Use of Definitions and Usage Samples by Four Terminology Banks

	For an Entry Term in Language 1		
	Definition L1		Usage Sample L1
Terminology Banks			
Montreal	x	or	x
EURODICAUTOM	x	or	x
NORMATERM (of the Association Francaise de Normalisation)	x		
CETA (Chinese-English Translation Assistance Group)	x		

In the banks using definitions, the justification is that a definition is necessary to identify the ideas characteristic of a certain term or field to insure information validity; other banks feel that the context of a term is more valuable to the translator in that it may contain important terms which are usually used with the entry term. In addition, the term is presented as it functions in a sentence. Guidelines for contexts were established at the Canadian Colloquium for Terminological Research [cf., Can3].

Three types of context are defined:

1. definitive context: Without the rigor of a lexical definition, it informs us about the precise semantic content of the term. Ex.: Weaving is a method of producing cloth by interlacing two or more sets of yarns... at right angles to each other. (Weaving)

2. explanatory context: It informs us about the nature of one aspect of the term, without defining it. Ex.: Each warp yarn must run straight from cloth beam to warp beam without being crossed with any other yarn; this is essential for raising and lowering the different harnesses to form a shed. (shed)

3. associative context: It just links the term to a particular subject area or to a group of related terms, as in a list, or purely functional situation. Ex.: Weaving is a method of producing cloth by interlacing two or more sets of yarns... at right angles to each other. (yarn)

It may be possible to streamline definitions to be more relevant for terminology banks. They may be more effective if, instead of presenting a formal stated equivalence, the definitions are a series of paraphrases or concrete examples to describe the entry succinctly. While a formal definition may add significantly to an entry, there are several other factors to be considered:

1. Definitions may make for cumbersome entries (if the translator is retrieving directly at the terminal).

2. Sometimes other forms of information may be more useful to the translator.

Recently several translators in the Washington, D.C. area were polled, and the kind of entry information they suggested did not include definitions or contexts. A sample of their suggested format is given in Figure 8.

Significant progress in this area will be achieved only when protocols of translation sessions are taken to determine cognitive effects of definitions and usage samples (See Part VI).

Figure 8. Entry Information Suggested by Translators in Washington, D.C.

мировоззрение (syn: политическое настроение)

world outlook
Weltanschauung
ideological outlook
ideology

обучение

training
Result of "обучение" is "образование."
"Обучение" is a means of "воспитание."

режим

regime (gov't)
procedure
system
control (e.g. воздушный режим: air traffic control)
schedule (e.g. строгий режим: strict schedule, as in strict schedule of soldiers' duties)

Matching contexts (or definitions) by translation

In a record, contexts can match each other or not. In the Montreal Bank the context of a term is the direct translation of the context of this term in another language. At the CCE Bank it is felt that contexts rather than being a direct translation of each other should be matched. This method has its own advantages. For example, the term "fretting corrosion" when associated in English context with the term "polish," is translated in French by "aviver" rather than "polir." More imfornation is thus gained by using the context matching method. However, this method requires more effort. After extracting the context of the word in the source language, the terminologist must locate a target language document where the same word with a similar context appears. [cf. Rei]

Field classification

Existing terminology banks have not adopted the conventional systems of classification. In their place they have developed classification systems corresponding to their own needs. The three classic systems, Dewey, Universal Decimal Classification, and the Library of Congress are based on systematic classification. They follow a breakdown of knowledge dating back to the nineteenth century. They impose classifications that do not match the scientific evolution that has occurred in the twentieth century. For instance in the Dewey system, "computer science" occupies a subsection with four decimals (621.3819) while "roof construction" has a complete section (695). Although the conventional systems offer many specific subcategories, many of them are useless to the translator.

Terminology banks on the other hand stress the need for a dynamic system, pragmatically based on contemporary technical developments. While the classical systems include only from 10 to 21 major fields, the American Translators Association has established the categories in which its members are specialized to be 82 while the Centre Nationale de la Recherche Scientifique (France) proposes 69.

The Montreal Bank developed a classified fields code with a three-level logical structure. Twenty-six classes form the skeleton of the classification and contain the basic fields of activity. Each of the 26 classes includes 26 divisions representing activities in each class; divisions are in turn subdivided into 26 sections. For example, "perfectionnement et formation du personnel" is described as belonging to categories on three levels:

A = Administration
F = Gestion du personnel
F = Perfectionnement et formation du personnel

It is coded as AFF. "The Montreal classification system's advantage over the standard codes already in existence [Universal Decimal Classification, Dewey, Library of Congress]is that it is designed exclusively for terminological utilization, and has all the flexibility necessary for that purpose."

[cf., Mon2] This mode of classification affords a theoretical possibility of 17,500 vocabulary sections.

The classification system of the CCE Bank has a two-coded category level [cf. Com]. A third level is present but only labelled, not coded. Thus, only general glossaries belonging to level one and two can be retrieved. To this third level, new fields may be added. Like the Siemens system, the CCE major fields are geared to the specific needs of the organization.

In order to understand how these systems work, we have located in the following chart (Figure 9) the concept "electric light" in the systems connected with the CCE and Montreal banks and the International Standards Organizations (Universal Decimal System). Each of these classification systems is a hierarchical branching structure in which the first category is "all knowledge." A search for the category "electric lights" begins with "all knowledge," and the branches are scanned until "electric lights" is located in the hierarchy. To measure the degree of specificity of each classification system, we establish two categories. "Categories scanned" measures the specificity of the system as it relates "electric lights" to "all knowledge"; it is the sum of the categories scanned during the search for "electric lights." "Categories encompassed" measures the specificity of the system as it relates the category "electric lights" to its own referents; it is the sum of all categories on all levels of the hierarchy which branch from "electric lights." Taken together the figures for "categories scanned" and "categories encompassed" describe the specificity of a classification system and the position of "electric lights" within it. It appears that the number of categories scanned to locate "electric lights" is fairly similar in these three systems: CCE = 59, Montreal = 41, UDC = 51. However, the number of categories specified under "electric lights" is quite different: CCE = 0, Montreal = 7, UDC = 232. Clearly, the Universal Decimal System includes a finer subdivision of categories than the other two systems.

The above systems all have in common a certain amount of rigidity; they are tree structures in which a node has only one predecessor. This is very limiting because there are in in fact many fields that can be the subfield of more than one field and thus cannot fit easily into the tree structure. For example, the field of publicity could be under "commerce" but also under "building industry." To solve this problem a classification system has been proposed by Robert Archimbaud of the Canadian Terminology Bank. [cf., Arc]. This system is characterized by:

1. dynamic evolution: new fields can be added, others deleted, in order to follow the evolution of the 20th century sciences,

2. a greater number of major fields (46),

3. possibility to connect a specific field to any major fields,

4. the possibility to subdivide fields into as many sublevels as desired,

5. field definitions based on concepts.

The fundamental idea behind this system is that a field is not fixed, located at a precise level of a well defined part of a knowledge tree, but is rather dynamic, and can be located in differnet places at different levels to correspond to the changing and expanding universe of knowledge.

Figure 9. Three Classification Systems

Organization Name	Classification System Name	Status of System	Location of Concept "Electric Lights"			Categories Scanned to Find "Electric Lights"	Categories Encompassed by "Electric Lights"
			Level	#Categories	Relevant Category		
Terminology Office, Translation Division, Commission of European Communities	System of Classification	Operational	0	1	All knowledge	59	0
			1	45	Electrical engineering & energy		
			2	13	Lighting (primary category)		
Terminology Bank, University of Montreal	Fields Classification	Defunct	0	1	All knowledge	41	7
			1	22	Electricity		
			2	18	Electric lighting (primary category)		
			3	7	All categories (are lamps,....)		
International Standards Organization	Universal Decimal Classification	Recommended	0	1	All knowledge	51	232
			1	10	Applied science, medicine,technology		
			2	10	Engineering		
			3	10	Energy		
			4	10	Electrical engineering,electrical industry		
			5	10	Electric lighting, equipment, lamps, primary category)		
			6	6	All categories (incandesent, lamps,....)		
			7	18	All categories (manufacture of incandesent lamps,....)		
			8	47	All categories (support,....)		
			9	97	All categories (wires for current leads,....)		
			10	50	All categories (removal of gas from current leads,....)		37

To implement this idea in a system, major fields and specific fields are defined. Although the number of fields is unlimited, there currently exist 66 major fields to which the specific fields are connected through a network structure of intermediate fields which themselves are either major or specific fields. This organization can be represented as in figure 10.

To fit the different possible subdivisions of a given field, there can be as many intermediate fields as desired. Since it is a network structure, different paths are possible to get to one field. For instance, the field "automobile" can be accessed either through "véhicules à moteur" and "mécanique," or through the "assurances" field.

Each record in the bank contains the different fields to which it belongs and through which it can be accessed. This system seems very flexible and appears to be an innovation among the classification systems used in terminology banks. For a complete description of it, see Archimbaud's paper [cf., Arc].

Organization of the Data Base of the CCE System, EURODICAUTOM

Since most of the existent terminology data bases are similarly organized, we will describe only the one for the EURODICAUTOM system of the CCE.

The EURODICAUTOM data base is composed of three types of files. Its organization is diagrammed in Figure 11. The main one, called the text file, contains the records. Each record is identified by a number which is used to retrieve it by the IBM ISAM accessing method. [1] This identification number is internal to the system and is different from the identification number used by the terminologist. [2] This implies the existence of a second type of file, a cross file which does the mapping between the BETYNI code and the internal code. The third type of file is an inverted file which contains character strings (terms) alphabetically ordered. To each character string, or term, is associated a language code and a list of internal keys referring to the main file. Each of them points to a record where this term appears either as the term or as an entry in the term. Each list of internal keys is left to the creation order.

Each record in the text file has a variable length which allows updating of its content. This feature is important because updating a record does not imply the creation of a new record but only a modification of the existing one. However, some banks, such as the Terminology Bank of Canada, have a fixed length record.

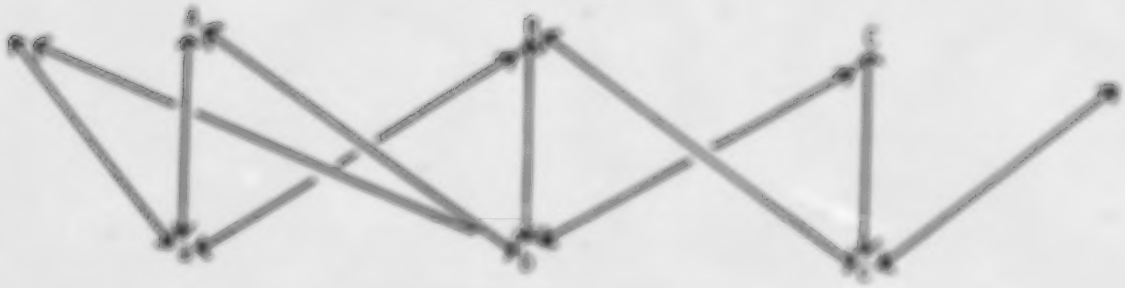
1/ISAM (Indexed Sequential Access Method) used with IBM 370. With ISAM files the records are grouped to fit onto the physical disk.)

2/The terminologist code is BETYNI: BE = Bureau émetteur (source office); TY = type of origin (for example, the field); NI is for identification number.

Figure 33. Proposed Classification System for the Canadian Terminology Bank

Architectural Systems: Major Fields = ABC
Specific Fields = abc

Major Fields



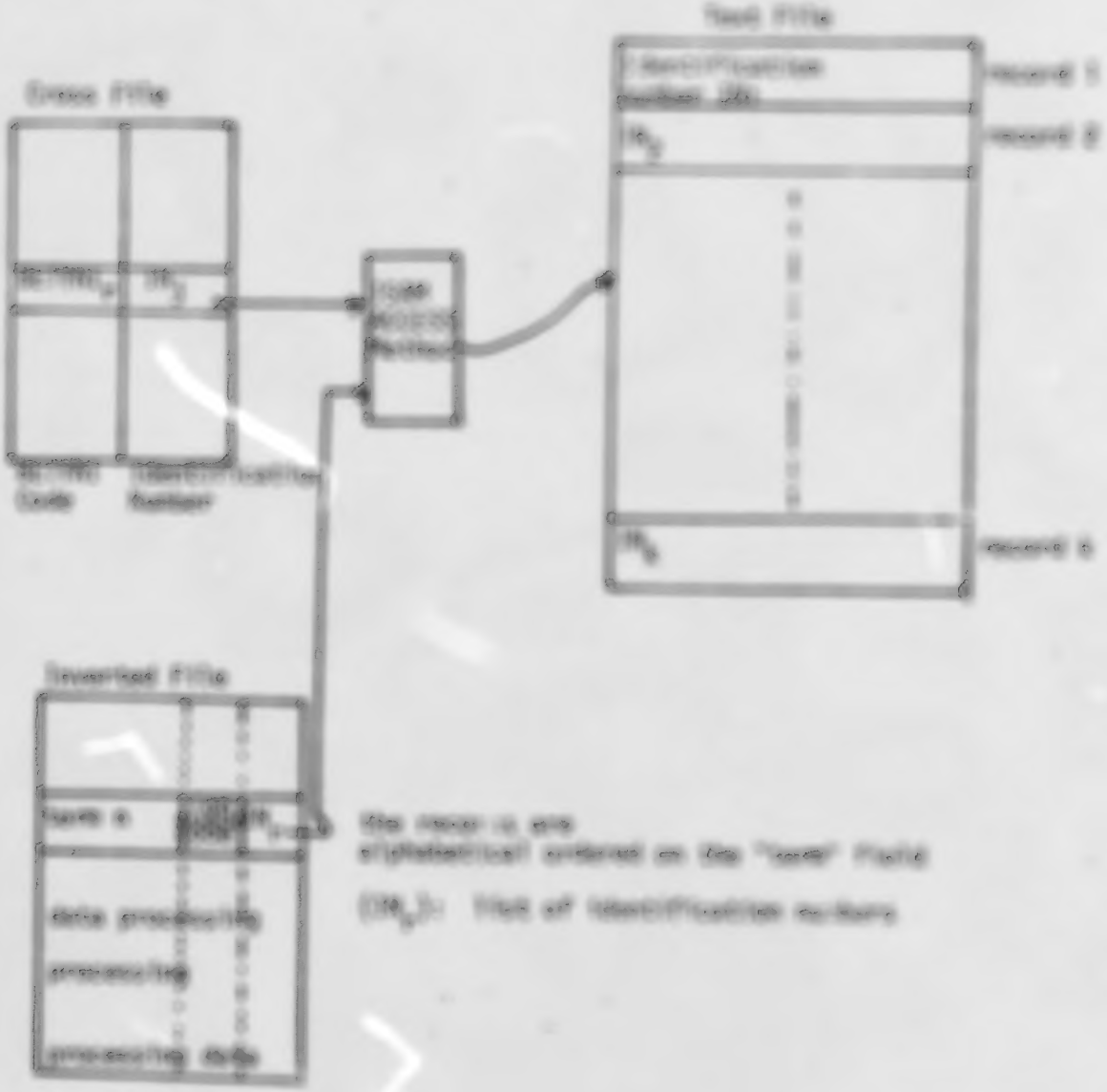
Specific Fields

Classical tree structure systems: Major Fields = ABC
Specific Fields = $a_1a_2a_3b_1b_2b_3c_1c_2c_3$

Major Fields

Specific Fields

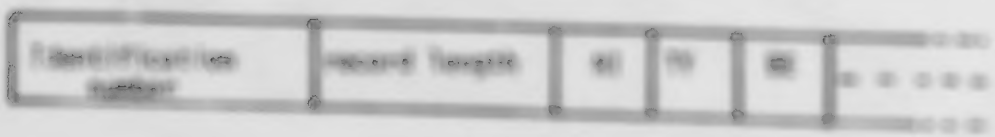
Figure III. [illegible] Data Base Organization



The records are alphabetical ordered on the "words" Fields

(ID_n): list of identification numbers

Structure of a Record



Comments on the EURODICAUTOM data base

The access to a fiche is either through its BETYNI code or through a term which is contained in it. The BETYNI access is used mainly to update the text file.

The adding of a new record to the text file implies the generation of an identification number by the data base management system, the allocation of a record in the text file, a new entry in the cross file, and the updating of the inverted file. All the relevant terms belonging to this new record are access terms. If they already belong to the inverted file we add to each identification number list the identification number of the new record. If a term is new, a new entry is created in the inverted file.

The structure of the inverted file allows direct access to the identification number of all the records which contain a given term. The inverted file has many entries because an entry may be either a key word (important constituent) of a context or a key word of a term. If we ask for the set of records containing several key words, the system does the intersection of the identification number lists associated to each key word and accesses the identification numbers which belong to the intersection in the text file.

Computer characteristics

The text file of EURODICAUTOM takes 40 million bytes, and the cross file takes 2.5 million bytes. The average size of a record is 300 bytes. This system runs on an IBM computer, using the OS MVT. The similar Siemens system running on BS 1000 (Siemens computer) takes 55 million bytes on disk with the same average size of a record of 300 bytes. The number of terms for some of the banks is: CCE, over 600,000; Quebec, 163,000; Canadian Terminology Bank, over 1,233,000; Siemens 2,000,000; Dresden 320,000.

Query Language

The query language allows the terminologist or the user to communicate with the data base. We can distinguish two kinds of communication: the creation and updating language and the interrogation language.

The creation and updating language allows the user to modify the contents of the data base. This language is used in the batch mode in all terminology banks and in the on-line access mode only in the former Montreal Bank and the Carnegie-Mellon University Bank. The interrogation language is used in all banks in the batch and direct access modes.

Very often, the fact of using one uniterm as an entry leads to a long list of records retrieved for a query, especially if there are several access fields to which the term belongs. To reduce the number of retrieved records, several steps can be taken. For example, the CCE Bank allows a list of terms to be queried simultaneously; only the records which contain all these entries are selected. Montreal added another parameter which specifies in which classification field an entry is considered. The pair (term, classification field)

is then used to select records. In the Normaterm Bank the query can be more selective since a full interrogation language using Boolean operators is available. [cf., Lau2] Thus, a request can, for example, be formulated in this way:

(Term = paper or feuille) and (field = électricite or électronique)

The point stressed here is that a good interrogation language should allow the user to give as much information as he can. During the process of translation, the information to be transferred into another language is not represented by the character-string of the term itself, but by the concept it represents, the fields in which it is used, its context, and many other things which are not easy to formulate. The translator is aware of all this and would like to communicate as much information as possible to the system. This is a difficult problem which is discussed further in Part VI. An example of commands useful for a query language is given in Figure 12.

Figure 12. A Set of Commands Useful for the Query Language [cf., Gre1]

Command	Definition
HELP:	to describe the existing functions
CREATE:	to enter a term. Ths user is prompted for the following: term: language: field: equivalents: grammatical category: CREATE then allows the user to select among several other categories. The question: Select: usage sample definition note synonyms may be answered by specifying any of these categories in any order. Once this has been done, the term is stored in the bank and entries are made automatically for the equivalent terms.
EDIT:	to update existing entries
TARGET:	to find term equivalents
RETRIEVE:	to display entire terms
DELETE:	to remove terms from the bank
RECORD:	to record interactions between the user and the system
Hard copy facilities:	to print out on a XEROX graphics printer: (a) individual terms: the TERM HARDCOPY command (b) dictionaries: the DICTIONARY HARDCOPY command (c) term names only: the LIST CONTENTS command

Making of Dictionaries [cf., Mac]

A terminology bank allows the printing of technical dictionaries in a very efficient way. For instance, at Siemens, the selection of about 10,000 entries from a total fund of over 200,000 and the simultaneous generation of synonym and inversion entries takes about 15 minutes, the sorting takes 8 minutes, and the typesetting program takes 6 minutes. Thus only a half-hour is necessary to print out an entire dictionary.

However, there are still some problems in the making of a dictionary. The main ones are selection, and synonyms.

Selection

Field classifications attached to each entry are the main criteria for the selection of expressions for a dictionary. Thus the more field sub-divisions are coded, the more selections of glossaries are possible. Other criteria are feasible, such as date of entry, author, grammatical characteristic, etc, but field classification is essential.

Another problem occurs when an expression in the source language is not matched by an exact equivalent but by a temporary expression or a circumlocution. Although helpful to the translator, these expressions should not appear as a search term in the target language. A selection of quality of entry is necessary when these temporary expressions (as opposed to clear and accepted expressions) are to be stored. This problem is solved automatically by a formal identifier in the program.

Finally, a dictionary will contain only a certain number of languages. A selection is therefore necessary which delivers only entries which contain information in the desired language. This suggests the possibility of a further selection within the entry itself: program parameters can determine what additional information, such as synonyms, source notation, definitions etc., should be included or suppressed.

Inversions [cf., Mac]

"A second, similar problem arises with expressions composed of more than one word. Compound expressions are especially frequent in technical fields. These multi-word expressions are usually read and stored according to the natural word order of their components. [e.g. 'bistabile oscillator': the adjective comes before the noun.] In the target language of the printed dictionary this word order is retained. In the source language, however, the needed information [the translation of this compound expression] is found not only under the first word [here the adjective], but also under any following words [here the noun], insofar as this is sufficiently important.

These key words receive a 'tag' or notation when they are collected, and this allows the program to generate inversion entries. These inversions are separate dictionary entries, in which the keyword is placed at the beginning of the entry and the rest of the expression is added after a comma. [Thus, the term 'bistabile oscillator' generates the inversion entry 'oscillator, bistabile' and is bound under both 'B' and 'O'.]

The tagged keywords can also be retrieved for eventual sorting and arrangement in the dictionary when they are part of a composite expression or compound word. For example, 'circuit, NAND-' goes with 'NAND-circuit' and 'camera, from ____ to lab' goes with 'from camera to lab.' This latter expression appears in the English part of the dictionary only in the first inverted form, since it is senseless to enter it under the initial word, 'from.'" [cf., Mac]

Synonyms

"Two other problems appear in connection with data sorting. The first is the question of synonyms, which, as already mentioned, are read together with the basic designation of a concept in a stem entry. Whereas in the target language(s) of a dictionary all available synonyms may simply be given one after the other, in the source language available synonyms must be placed in quite different places, according to their spelling. They should either give complete information on the stem entry, or simply give a reference to where the user can find the necessary information. To meet this requirement, all synonym entries or references are generated at the stem entry in the source language of a dictionary before the actual sorting per program." [cf., Mac]

Terminological digests [cf., Lip]

A terminological digest is a list of terms extracted from a main dictionary in the order of textual occurrence instead of alphabetical occurrence, i.e., the desired terminological framework of a text to be translated. The interest in such digests has been shown previously, and now we are concerned with their efficient production. Figure 1 of Appendix II represents an example of a text portion to be translated; Figure 2 shows the automatically-produced terminological digest of this text in the order in which the desired terms occur, and Figure 3 shows the terminological digest in alphabetic order [cf., Lip]

Such digests can be produced in two ways:

1. manually: the list of the terms to translate is established by the translator. He can then enter it either in the batch mode (on cards) or by direct access via terminal. The system will access the terminology bank and print out the terms with their translation.

2. automatically: the list of the terms is established automatically by the computer. This mechanism is described in Part VI. Here the system accesses the terminology bank and prints out the digest. The printout can be in textual or alphabetical order, the sorting being done by the system itself [cf., Appendix II]

Other Procedures

Conversation updating

The conversational updating procedure allows the user to add, modify, or suppress records in the data bank directly through a terminal. It replaces the cumbersome method of batch processing. However, it has its drawbacks. File protection has to be enforced to avoid uncontrolled modification of the data base, and preserve its quality. When several users are allowed to access the data base, complications can arise when more than one is modifying the same record. The quality of the data is then endangered.

Different solutions exist for these problems at the data base level. The Montreal Bank, which has this feature, prevents system access while the updating occurs. Other banks feel that the addition of this feature is not cost effective. Thus, they update the bank in a batch mode when other users are not accessing it.

Truncation function

The truncation function allows for various morphological variations of a term, such as compounds, plurals, etc., to be accessed when a fixed number of initial characters of this word is typed. For example the character string "consider" should retrieve the following terms: "considering," "considered," "consider," "considers," etc. The German expression "kenntnis...geben," should retrieve among possible other answers "kenntnis...geben" and "...gibt...kenntnis."

The fixed number of characters (substring) to be typed in order to access all morphological variations is language dependent and is also a function of the length of the term. For example, a term of seven characters might be accessed by a character substring constituted by its initial four characters, but a shorter term may be accessed by less than four character to a minimum of two.

An interesting feature has been developed in the EURODICAUTOM System. Figure 13 gives an example of truncation values for a substring.

Figure 13. Truncation Values for a Substring (EURCDICAUTOM)

Language	Length of the interrogation term	Truncation	Linguistic origin				
			VE1	MC1	VE2	MC2	AUT
any language		-4	200	100	150	50	0
		-3	500	400	450	350	200
		-2	600	500	550	450	300
		-1	700	600	550	550	400
		0	1000	900	950	850	700
		+1	600	500	550	450	300
		+2	400	300	350	250	200
		+3	300	200	150	00	80
		+4	100	50	80	25	10
		+5	10	5	10	5	0
		+6	10	5	0	0	0

Note: '-' means that the answer term (ex. ETALONNE) contains the interrogation term (ex. ETALON)

'+' means that the interrogation term (ex. GONADES) contains the answer term (ex. GONADE)

Figure 13. Continued

Language	Length of the interrogation term	Linguistic origin / Truncation	VE 1	MC 1	VE 2	MC 2	AUT
English. French, Italian	2-3	-1	100	99	98	97	96
		0	105	104	103	102	101
		Others	0	0	0	0	0
	4-5-6	-1	100	99	98	97	96
		0	105	104	103	102	101
		+1	100	99	98	97	96
		Others	0	0	0	0	0
	7	-2	100	99	98	97	96
		-1	100	99	98	97	96
		0	105	104	103	102	101
		+1	100	99	98	97	96
		+2	100	99	98	97	96
		Others	0	0	0	0	0
German, Danish, Dutch	2-3	-2	100	99	98	97	96
		-1	100	99	98	97	96
		0	105	104	103	102	101
		Others	0	0	0	0	0
	4-7	-2	100	99	98	97	96
		-1	100	99	98	97	96
		0	105	104	103	102	101
		+1	100	99	98	97	96
		Others	0	0	0	0	0
	8	-3	100	99	98	97	96
		-2	100	99	98	97	96
		-1	100	99	98	97	96
		0	105	104	103	102	101
		+1	100	99	98	97	96
		+2	100	99	98	97	96
		Others	0	0	0	0	0

cf., Pral

Searching function (CCE Bank)

The strategy adopted by the CCE Bank is to order answers to queries with respect to the quality. This searching function allows the best answer to a translator's query to appear first followed by other possible answers in order of decreasing importance. This function takes into account the language, the truncation function, the fields, etc., with different weights for each parameter [cf., Fra]

If several key words are typed, contexts are obtained in which they appear together. The CCE computer scientists plan to develop a way to access through the term and the field of interest. The decreasing order of importance of the records should allow the user to find out quickly, without having to read the whole list, if he has the desired answer. The following is the system used for the EURODICAUTOM system.

EURODICAUTOM Classification Function [cf., Fra]

When a request is input, to each entity of the bank is associated a function $f(x_1,...x_n,y_1,...y_m)$, and entities are sorted out in the decreasing order with respect to this function.

$x_1,...x_n$ variables describing the linguistic and qualitative characteristics of the entity

$y_1,...y_m$ local characteristic variables

x_1 linguistic origin of each uniterm i belonging to the request (VE1: the term is composed of one word,
MC1: the key-word expression is composed of of one word,
VE2: a key-word extracted from a term,
MC2: a key-word extracted from a multiword key-word expression
AUT: a word extracted from a sentence)

(Thus a first classification appears: if a word belonging to the request is VE1 in an entity (or record) and is AUT in another entity, priority will be given to the entity in which this word is VE1.)

x_{2i}: indicates the existence of each uniterm i, belonging to the request, in the answer entities (empty words are neglected in the request)

Example: Request: BIT (DE) PARITE

For an equal configuration of other variables, the first answers will contain BIT (DE) PARITE, and then entities containing BIT or PARITE,

x_3: characterizes the identity of the classification field of the request with the classification field of the answer entity,

x_4: characterizes the value of the reliability code of the answer-entity,

x_{5i}: characterizes the matching (truncation) in a given language between each uniterm i and the term in the answer-entity; the function decreases proportionally to the length of truncation.

Example: Request: TOLE

All other variables being equal, the output order will be:

TOLE
TOLES
TOLERIE
TOLERANCE

y_1: characterizes the importance of the bureau which has input the entity in the bank.

y_2: characterizes the importance of the "TYPE" (the "TYPE" indicates if the entity belongs to a glossary or not).

y_3: describes the existence of several target languages: the user starting from French to German, prefers to receive first French - German before French - English, even if this last entity appears to be "better" than the other.

The f function has to be simple, and the value of the variables are given in arrays. Thus, these values can be modified to improve the quality of the search.

In the batch mode the function is:

$$f = \sum_i x_{1i} \cdot x_{2i} + x_3 + x_4 + y_1 + y_2$$

with:

x_{5i}: are mixed with x_{1i} (f table)

x_{1i}: value common all languages (f table)

x_{2i}: $x_{2i} = 0$ if no answer

$x_{2i} = 1$ if no answer

x_3: $x_3 = 100$ if the classification field required belongs to the answer

$x_3 = 0$ otherwise

x_4: $x_4 = 75$ if the reliability code is lower or equal to the code of the answer

$x_4 = 0$ otherwise

y_1: $y_1 = 80$ if the bureau is the required one

$y_1 = 0$ otherwise

y_2: $y_2 = 90$ if the type of the answer is the required one

$y_2 = 0$ otherwise

V. EXCHANGE PROBLEMS IN NETWORKS OF BANKS

Individual companies or institutions across the world are not in a position to acquire all the specialized terminology in existence or to keep terminology stock up to date. [cf. Uni] Thus, there is urgent need for international cooperation in the field of technical terminology between industries, governments, and public as well as private institutions. To deal with this requirement, the International Information Centre for Terminology (Infoterm) was established in 1971, sponsored by the UNISIST Program (a branch of UNESCO). Infoterm is active in the field of international terminological standardization, and promotes cooperation among terminology banks. It works in liason with the Technical Committee 37 (TC 37), "Terminology (principles and coordination)," of the International Organization for Standardization (ISO). One of the sub-groups of TC 37 is Working Group 4 (WG4), "Computational Aids in Terminology and Lexicography," organized to establish an international terminological format (data unit) to facilitate exchange of terms among banks. It has published many useful reports [cf., Feli-Feli]

Terminological cooperation between international organizations necessitates standardization of terms and definitions. Hence the need for standardized symbols (to be used in definitions) and for standardized terms. Standardized symbols have been established by Infoterm, to be used in classified multilingual defining vocabularies. The same symbols may also be used in alphabetical vocabularies and in terminological papers. The selection of the recommended symbols [cf., Iso] was prompted by known usage and by the insights acquired in lexicography, linguistics, and symbolic logic.

Networks of Banks

The joint operation of data bases can take place in several ways. The most simple one is by exchange of magnetic tapes on which the data is recorded. Bank A will record its terms on tapes and send them to Bank B which will read them in. As simple as it looks, this solution raises many compatibility problems cf., Iso4

Data Structure Problems

The information concerning an entry, described in the previous section, is organized in a data structure called a record. A record is a set of fields, each field containing an item of information (for instance, one field contains a term, another its definition, etc.). The number of fields used for an entry depends directly on the number of attributes linked to it and is a characteristic of the data base.

Since the exchange between two banks involves the matching of two data structures, if these data structures are not identical (i.e., same fields of variable length), a loss of information will necessarily occur. For instance, a field which exists in one data base may not exist in another. Or, the length of the receiver field may be smaller than the length of the sender field. Hence, it is necessary to have variable size records and as many fields to characterize as entry as possible. [cf., Fao]

Linguistic Problems [cf., Lee]

It is not enough to have the same fields at the record level. Matching fie
from different banks should have the same meaning. The most striking examp
is certainly the field of "field classification." Each existing terminology
bank has its own classification field (the CCE Bank, Siemens, etc.). This
that every time terms are exchanged they must undergo conversion. It is ob
that this conversion procedure loses information and introduces inaccuracy.
international field classification system has not yet been established. Wu
Felber (from Infoterm) proposes an umbrella classification, defining at lea
the major subject fields and leaving the classification of the subfields to
the individual banks.

Exchanging contexts also presents problems. In the CCE Bank, each term is
associated with a context in each language and the contexts are translations
of each other. In the former Montreal Bank, however, separate contexts are
from each language.

Redundancy Problems

Adding terms to a bank should be controlled to avoid duplication. It is usu
easy to deal with duplication, except when several entities, coming from di
sources, are merged together to give one unique entity whose origin cannot b
defined precisely. For example, when one French-English record defining an
is input and another English-German record defining the same entry exists a
in the data base, the computer system should make only one record for these
This requires a search for the equivalents in the two languages, French and
English. Then the terminologist must decide what information to keep for th
twice-mentioned English term.

When adding a term to a bank, the problem of matching senses may arise. For
French and German terms which match the same English equivalent may not matc
other. Here again the terminologist has to intervene (see Part IV).

Computer Networks

A network of computers is a set of interconnected data bases. The connectio
via phone lines. This interconnection means that one computer can request a
from another computer belonging to the network and receive an answer in seco
Some large computer networks exist; for example, ARPA in the United States a
Cyclades in France. Problems are encountered in the establishment of such
networks. [cf., Lee] These involve compatability of different types of comp
communication protocols, query languages, data structures, and character cod
(diacritics, umlauts, etc.).

Other problems involve locating terms which are distributed among the numerou
of the computers belonging to the network; for instance, how can one find th
location of one given term? What if one term exists in several copies in di
places? Research is being done in this area. One solution to the problem i
have each computer center specialize in one (or several) fields. For instanc
center A will specialize in computer science, center B in electronics, and c
C in management. This will avoid duplication of data (one term in a given f

will exist only in one place), simplify updating of the whole term, yet cover each subject in depth.

Infoterm has proposed the creation of a universal model for a terminology term which would define a common language interface among existing terms. If each existing term is able to express its data structure in terms of the universal model, communication will then be achieved. Thus there would be no need for each computer to translate its data structure into the data structure of the receiving computer. Computer A may communicate with computer B without knowing the data structure of B. Processor A will transform its information into the universal model and processor B will reformulate this into its own model. Thus a terminology term need only express its content in terms of the universal model.

VI. FUTURE DEVELOPMENTS

This chapter includes a short review of issues in (1) data base organization, (2) translation research, (3) interactive translation systems, and (4) man-computer communication.

Issues in Data Base Organization

Issues in the data base organization of terminology banks include the possibility of relating terms to their equivalents through a concept, rather than directly.

The concept model [cf. Sec]

The conventional mode of connecting terms in data bases is to link term to term. We will call this the "direct link method." However, the Infoterm Data Bank, a recent document [cf. Sec 6.8

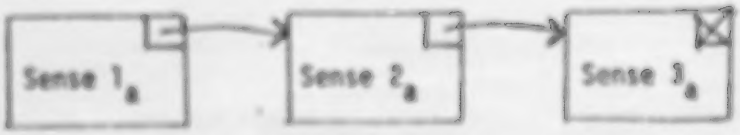
> The new Fichen [record] will be single-concept fichen: users will proceed from the concept to the term rather than from one term to another. This will make it possible to group related terms (synonyms, abbreviations, spelling variants, and so on). All new entries related to concepts already in the bank will be additions to the existing concept fiche rather than new fiches. This is the only method of setting up a relational, multilingual file, in which each concept is presented with its terms in each of the languages involved.
>
> Since the fiches are organized by concept, there could be several under one terminology unit (English or French), in cases where the terminology unit has homographs that refer to several different concepts. For example, for the English word "pipe," there would be a fiche on a "tube with a small bowl at one end, in which tobacco and, to smoke" [Webster's], since they are two distinctly different concepts.

In considering the TAUSY system, the "concept" and the "direct link" method of linking terms were discussed and some conclusions drawn. These are discussed below [cf. Page]

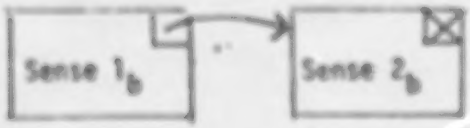
A term is a character-string to which is attached one or several senses [meanings]. Independently of how the senses are formed, the crucial problem of representing the equivalence with the senses of other terms arises. We are interested in connecting terms belonging to different languages which have the same sense.

Figure 1. Terms and their Senses

Language A: ⟨term a⟩



Language B: ⟨term b⟩



In Figure 1, ⟨term a⟩ and ⟨term b⟩ are equivalent if they have one sense in common. They are then a straightforward translation of each other. This can get more complicated if for a given term and sense there is no equivalent sense in another language. In that case, the translation of the term is made easier if some comments are made. This is shown in the following example:

French: table de salon

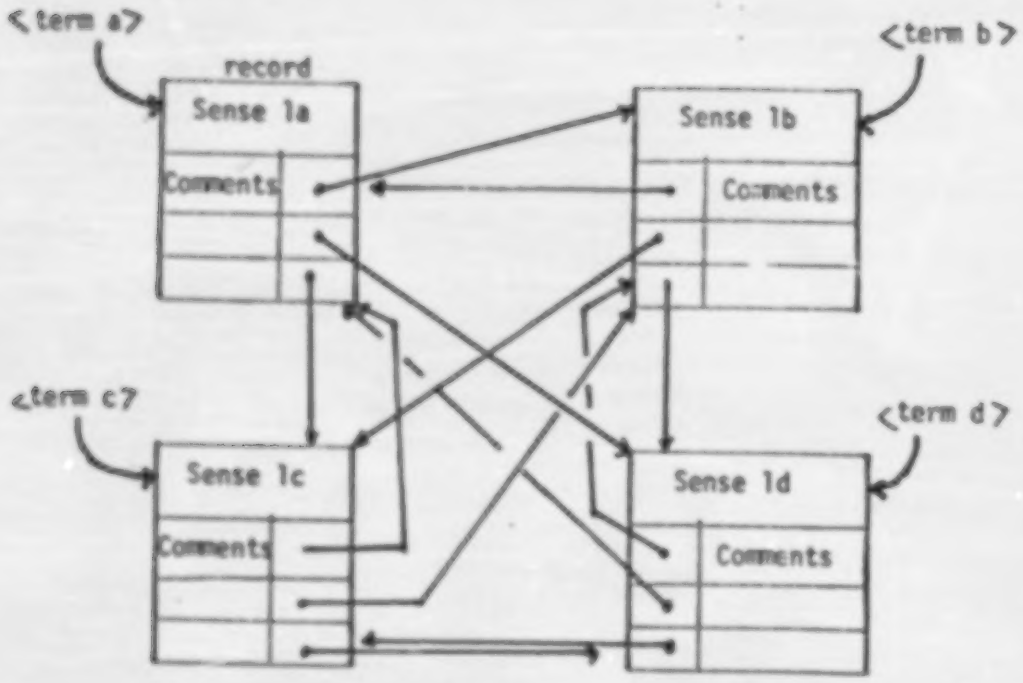
sense: table se trouvant dans un salon

note: this term does not exist in English. It includes "coffee table," "cocktail table," "end table," and other living room tables.

English: none — no equivalent

The exact same problem appears when two terms have nearly the same meaning. Their connection needs not only a pointer to each other but also some comments to characterize it. If more languages are added to the bank, more connections should be established. This is illustrated by the following figure in which four languages are considered.

Figure 2. Interconnection of Equivalent Senses



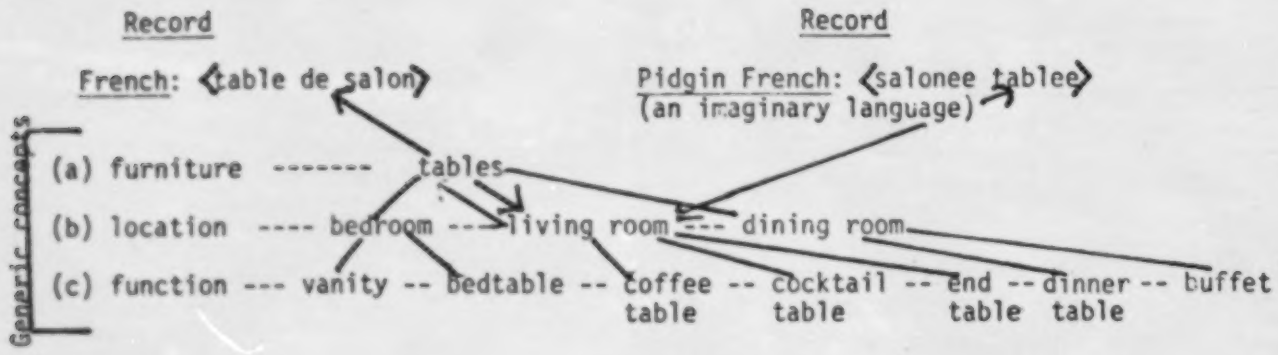
Each time a new language is added, this implies the modification of existing records by the addition of the pair: pointer, comments. From a computer scientist's viewpoint, this modification of existing components of the bank is something to avoid since it involves extra mechanisms (updating, more memory space, etc.); also, the number of pointers between terms grows rapidly with the number of languages.

Is there a way that a terminology bank can grow without affecting existing components?

Indirect connections between terms

One point of view is that terms are used only to designate concepts which are mental representations of material or immaterial individual objects, qualities, relations, etc... . To handle the relationship through a concept, generic concepts are formed for the obvious categories and the term is related only to its concept. The fact that the concepts are generic means that they are organized in a hierarchical way (for different relations of concepts see Felber's paper [cf., Fel]. Using generic concepts, we get the following structure for the term "table de salon":

Figure 3. Concept Structure for "Table de Salon"


Record
French: ⟨table de salon⟩
Record
Pidgin French: ⟨salonee tablee⟩
(an imaginary language)
Generic concepts
(a) furniture ------- tables
(b) location ---- bedroom --- living room --- dining room
(c) function --- vanity -- bedtable -- coffee table -- cocktail table -- end table -- dinner table -- buffet

Several remarks can be made on the above figure:

1. Only one pointer is required for each term's record. This pointer may have some comments attached to it but it is independent of all other languages. Thus, the adding of a new language does not affect the structure of existing records,

2. The number of pointers in between terms is reduced, compared to the direct link structure,

3. The descriptive information in each record is reduced since it appears in the generic concepts part. There is then less redundant information.

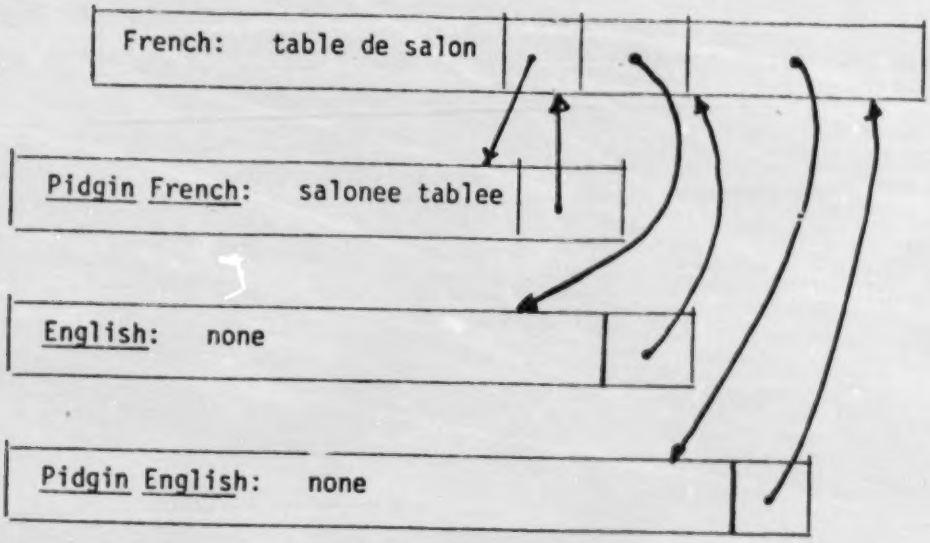
4. A term with no equivalent is related to target terms which are not its exact equivalents.

However, there are some great disadvantages in using indirect links through generic concepts. For instance, the intermediate concept system used for "table de salon" classifies furniture according to types (level a), location (level b), function (level c). This categorization is based on French concepts. How can the American concept "TV table" be incorporated into this system? This table is not confined to any room location; its location is determined by the presence of a TV.

Also, the adding of a new language or of a new term may bring new concepts which don't fit the concept structure. There is always the question of whether new intermediate concepts are needed or not, and whether a concept has now become redundant and should be removed. This all requires considerable processing.

An interesting variation of the above scheme is the use of a language as an intermediate level between terms. Instead of generic concepts, the intermediate level can be a language included in the data base. This allows an avoidance of the concept structure and its related problems. It also does not require extra links to a concept system. If for instance we take the French language as the intermediate language, we get the following structure:

Figure 4. Intermediary Language Links for "Table de Salon"; French Links Equivalents in Pidgin French, English, and Pidgin English



However, using a language as an intermediary for the translation of other languages also presents problems. For instance, let us designate English (not French) as the intermediate language; this is the obvious choice of an American bank where the target or source terms will usually be in English. How can we translate the French "table de salon" into "salonee tablee" in Pidgin French? We must go through the intermediary language, English, in which no equivalent term exists. Clearly some notation will have to be added in some language. Also, the input of two equivalent terms which don't belong to the intermediate language implies the input of the equivalent term belonging to this intermediate language. That makes the input of new terms cumbersome.

In conclusion the idea of an intermediate level is theoretically satisfying but is very difficult to implement. For this reason only the direct link method is used in existing terminology banks.

Precision in the request

One problem in the retrieval of existent terminology from data bases is the lack of precision in the formulation of the request by the user. We saw previously that one term in an entry can be connected to many records. The number of records which may be retrieved for a request can be reduced substantially if the request is well formulated. The "noise" in the request depends in some cases on the non-specification of term meaning and in other cases on spelling errors. Term meaning can be specified through the use of types of information on the term record. These types function as parameters which may characterize the meaning of a record during retrieval, if the system design permits. The types of information include the field of classification, the context, the dialect, and others judged to be relevant.

Spelling errors present a greater problem. The misspelling of a term is something which occurs quite often; especially in unfamiliar languages. However, a misspelled question should not lead to extreme behavior of the computer system, i.e., no answer or many answers having nothing to do with the question. The computer should tolerate lack of precision in the spelling and should be able to give intelligent answers. This "lack of precision" concept is studied in the IRIA research project called "Consultation de Donnees." Specific functions have been defined and implemented on special hardware. [cf., Iri] One type of solution for this problem is to define degree of precision. Answers are then given in decreasing order of precision; those more closely related to the request being presented first. This requires two things: a function to determine the degree of precision and a large look-up of the data base.

A consequence of implementing spelling-correction algorithms is a larger look-up of the data base since more data have to be read and evaluated by a precision function. Large look-ups of the data base will also result if the user can access through different types of information, or keys. In the ISAM method an inverted (or indexed) file is associated to each key. This substantially increases the memory size required to hold the data base since information exists in several copies. For a high number of keys (≥ 2) it is out of the question to have an equal number of indexed files. Thus, efficient searching functions have to be defined. A solution to this problem has been proposed by the development of searching functions at the disk controller level (IRIA).

Issues in Translation Research

Recommendations of the ALPAC report

The existing terminology banks have achieved enormous success in the areas of terminological classification and dictionary making. The banks of the Bundessprachenamt, Siemens, CCE, Quebec, and the Terminology Bank of Canada serve a vast group of translators efficiently, and have produced terminological principles and solved problems of computer system design.

Generally the problems of building a terminology bank have been attacked in relationship to linguistics and documentation, but only superficially in relationship to the process of translation. In these systems translators themselves do not interact on-line. Rather, specialist terminologists act as intermediaries between the data base and the translators. There is little research on the user's phase of the user-machine interface. In the first place, no studies are available about the process of translation itself. In addition, interactive translation systems have not been tested extensively and therefore, not much data on cost effectiveness and improved efficiency has been generated. However, see Mathias' "Machine-Aided Translation Experiment." [cf., MarF] It may be useful here to recall the recommendations of the APLAC report as a still valid guide for research in this neglected area. Six of the recommendations are listed on the following page.

Research is recommended to determine:

1. the cognitive process of translation,
2. ways to speed up this process,
3. delays in the over-all translation process, and means for eliminating them, both in journals and in individual items,
4. the relative speed and cost of various sorts of machine-aided translation,
5. ways to adapt existing mechanized editing and printing procedures to the process of translation,
6. production of adequate reference works for the translator, including the adoption of glossaries that now exist primarily for automatic dictionary look-up in machine translation.

In the following pages research efforts to implement the recommendations of the ALPAC report are presented.

Translation research at CMU

These recommendations have inspired research on the translation process at Carnegie-Mellon University. This research includes (1) an analysis of translation protocols, and (2) an analysis of machine-aided translations.

Systems directed to aid translations should be built on empirical data collected from translation protocols. Protocols of translators at work are taken to assess the cognitive routines used during this process and the type of aids which would be appropriate.

Protocol analysis is a technique whereby a person's behavior in accomplishing a task is minutely analyzed in order to derive a model of how s/he does it. Although it was used in psychological studies in the last century, it fell into disuse until revived at Carnegie-Mellon University during the last decade and has proved invaluable in the investigation of cognitive activity. This technique is being used to study what translators actually do during translation, rather than relying solely on preconceptions or textbook prescriptions.

Research was done by analyzing protocols taken at the beginning of the TARGET project at CMU. Some excerpts of protocols analyzed during the project are reported below.

Protocol 1

The following is the piece to be translated as broken down by the translator:

1 (je vous transmets ci-joint) 2 (la note d'information économique) 3 (que vous m'avez demandé) 4 (de vous établir) 5 (chaque mois) 6 (à l' occasion des Conseils Copperweld) 7 (vous voudrez bien me faire savoir) 8 (si ce type de document vous convient) 9 (ou dans quel sens vous désirez que nous vous le modifions le cas échéant)

The translator writes a rough translation first and then a more refined one. There appear to be two levels of ambiguity in the document: ambiguity level A is related to knowledge of the common vocabulary of a language; ambiguity level B is related to knowledge of its technical terminology. To ambiguity level A belong the phrases numbered 1, 3, 4, 5, 7, 8, 9. Their "disambiguation" depends on the translator's knowledge of the common vocabulary. To ambiguity level B belong phrases 1, 2, and 6. Their "disambiguation" depends on the translator's knowledge of the technical terminology.

The first draft of the translation contains many choices for the most appropriate translation.

phrase	level	choices
1 (je vous transmets ci-joint)	A/B	I am sending you attached to this Please find enclosed the
2 (la note d'information économique)	B	the note, the economic information
3 (que vous m'avez demandé)	A	that you have demanded asked
4 (de vous établir)	A	to establish or to set up for you set up compose
5 (chaque mois)	A	each month
6 (à l'occasion des Conseils Copperweld)	B	on the occasion or at the time of or for the Copperweld meeting or board of directors
7 (vous voudrez bien me faire savoir)	A	Kindly let me know Kindly inform me Kindly inform
8 (si ce type de document vous convient)	A	whether this document whether this kind of document whether this kind of memorandum whether this is the kind of memorandum you want
9 (ou dans quel sens vous désirez que nous vous le modifions le cas échéant)	A	in what direction you desire us to modify it as necessary or how you would like us to modify it or what modifications you might find necessary or what modifications you might want us to make

Since the choice of the best equivalent at ambiguity level A depends on the translator's general linguistic skill, he can by himself decide that "set up" is better than "establish," that "asked" is better than "demanded."

But there appears also to be a middle level of ambiguity for expressions which have one translation in the common vocabulary ("I am sending you attached to this") and another in a specialized field (business: "Please find enclosed"). This we call level A/B.

Finally level B of ambiguity refers to technical expressions. We have two of these. In phrase 2 the expression is disambiguated by looking further ahead to where the "note" is referred to as a "memorandum."

However, "disambiguation" is only the first step of translation. It leads to a literal target translation. The second step is from a literal to an idiomatic translation. In this case, the product of the first step displays a transformation from what in French is politeness to what in English is wordiness. In the next step, the translator must tranform the wordiness of "in what direction you decide us to modify it if necessary" to the idiomatic "what modifications you might want us to make."

For the above translation, no look-up of technical terms was needed. But this was not true for the translat on in Protocol 2, from a more difficult source text:

phrase	level	choices
1 (Consumos energeticos)	A	energetic consumptions energy consumptions
2 (Todos estos datos servirán para concluir)	B	all these data will serve to conclude (look up)
3 (sobre el diagrama de proceso más adeguado)	B	to come to a conclusion on the diagram graph (look up) process graph
4 (por la unidad de concentración a adoptar)	A	for the unit of concentration to be adopted
5 (donde se deberá tener muy en cuenta lo siguiente)	A	where one should keep in mind the following
6 (Si bien es cierto)	B	I wonder if (look up)
7 (se necessita lograr)	B	to choose (look up)
8 (un producto de Ley Fe)	B	a product of sixty-five Fe (look up)
9 (o mayor con porcentaje)	A	or better with a percent (%) of
10 (par tal razón deberían ser entregados)	A	For this reason For that reason they should be submitted
11 (ensayos exhaustivos)	A	to exhaustive tests
12 (de nodulización)	B	tests of pelletization (look up)

To ambiguity level A belong phrases 1, 4, 5, 9, 10, and 11; to ambiguity level B belong phrases 2, 3, 6, 7, 8, 12. In this passage a high percentage of the technical or semitechnical terms were looked up. The terminological work occupied 50% of the translating time. The other 50% of the translator's time was spent in improvements of style. This is an important point which the protocol makes absolutely explicit. There are two cognitive processes involved in this translation: one is a search for equivalents; the other is a search for the stylistically correct equivalent. We can infer that stylistic improvement time is a constant in all translating, while terminological looking up is a variable depending upon the difficulty of the text and the experience of the translator.

Three conclusions can be drawn from the above data:

1. First of all we should rule out the possibility of providing any aid for reducing the time spent on phrases of level A of ambiguity, since this is related to the translator's knowledge of the common vocabulary.

2. However, time spent on phrases of level A/B and B can definitely be reduced by having access to a good technical terminology bank. Particularly in the case of the word "pelletization" in the second translation, where no available dictionaries helped and some technical documentation for special new processes such as "pelletization" would have been useful.

While it is clear from these two translations that a terminology bank improves the quality of translation, it is not as clear exactly how much time-saving is involved. In the prototype TARGET system at CMU, a series of tests are being run to measure the time-saving effected by a terminology bank. Two pilot tests have been run so far. Each pilot test consisted of a translation using the system's retrieval program. Each translator translated the beginning of the same financial article ("L'intéressement dans l'entreprise," Bulletin hebdomadaire vol. 31, no. 43, Nov. 12, 1976, pp. 1-4) from French to English. The translators were observed by a member of the staff; their evaluations of the interactive system were recorded, as well as their time per word. The translators went through these tasks in the following order: extracting source terms from a document, querying the data base, jotting down target terms, transcribing the final choices for target terms.

Substantial time-saving produced by substituting retrieval from a terminology bank for dictionary lookup would be obtained here only if all the terms queried were in the bank and no supplemental dictionary lookups were necessary. As far as translation aid rendered by terminology banks goes, their contribution does not demand the utmost sophistication of computer technology, but rather a wealth of good terms in the bank.

Terminology banks may be augmented by the addition of new procedures to further reduce translation time. Translators spend a great deal of time writing and rewriting drafts of the target text. This task is as time consuming as the task of tracking new technical terms. It is clear that adding an editing and formatting facility to a terminology data base should not only improve the quality of translation but also its efficiency in terms of time saving and cost. With the addition of this facility, the translator may directly type the target text into the terminal as he composes it, and simultaneously retrieve unknown terms which will be displayed for him.

Issues in Interactive Translation Systems

In March 1976 the Federal Broadcast Information Service called a seminar on Aids to Translation (March 8-9, 1976) [cf., Jor]which concluded that "high quality MT is still impossible and will remain so until and adequate means is developed to represent the conceptual meaning of language texts." Developmental effort should be put on "a machine-aided translation system which would allow a human to interact (via CRT and a text editing system) with an on-line dictionary and other search, display, and formatting aids to produce high quality translation with time and effort much reduced over conventional modes of human translation." The idea was to use established computer capabilities to speed up and take over the routine parts of the translation effort.

This kind of system is proposed by Martin Kay and Erhard Lippman. Mr. E. Tanke, also [cf., Tan] has described computer procedures to simulate routine tasks in translation.

Tanke: artificial intelligence and machine aids to translation

Mr. Tanke believes that this type of system falls under the domain of the field of artificial intelligence. He comments that:

> The basic criteria which we wish to establish for [evaluating] the intellectual achievement of a computer are that (1) the solution which it finds [is] a novelty, in other words, the specific result must not be existent as a possible solution in the [memory] , and (2) that this novel solution is appropriate. Moreover, if the computer can perform any tasks that are believed to demand human intelligence, then we shall assume the computer to possess artificial intelligence.

The extraction of technical terms from a text can be automated. We will quote Mr. Tanke to describe the mechanism and the implementation of this technique:

> At present we [Siemens Co.] are developing a program which will be able to extract the technical terms from these machine-readable texts automatically and output text-synchronous word-lists without human intervention. The translator will then have no preliminary work to do on a text, such as underlining terms and so on, and no operator will be required to punch the queries. The translator will receive both the text-synchronous technical word-list and the text to be translated at the same time.

Such a task requires the following types of linguistic analysis:

1. The identification of grammatical classes of words: recognition of nouns, adjectives, verbs, etc. This is done by a program which operates on morphological rules and not by comparing the text word-by-word with an extensive vocabulary. This program is for the German language.

2. The stemming of words: reduction of words to their stems.

3. The analysis of syntactic structure. This is based on sentence patterns. During the process of analysis, multi-word terms (syntagmic structures) are located in the following manner:

[This process is used for the] definition of syntagmatic structures, which include compounds and idioms: since each subject field has its own typical syntagms, the words which belong together must be defined, along with the subject field or subject fields dealt with in the text. With homogeneous texts, such as the description of a piece of equipment, this is simple. Such texts can often be classified by their title. Special problems arise with those texts which touch on a number of different subjects, such as texts dealing with applications of EDP systems. In such cases one would start with a rough classification, obtain the first syntagms from this, further refine the classification using these syntagms and so on. As far as we can judge, the definition of syntagms will only be possible on the basis of statistical investigations. For this purpose, large volumes of texts are examined, the compounds and idioms are retained and stored in subject-related lists in the computer. In practical applications the text components, e.g., verbs and nouns, will then be compared with these lists using concordance programs. Where a match is found the text will be rechecked to see if further elements of the possible syntagm occur in the vicinity of the concordance. If this is the case these groups of words will be marked in the text as belonging together.

To determine the subject fields we shall use matrices or lists in which the technical terms for any given subject are listed. These will form the basis of the comparison and substitution process.

Before technical terms may be extracted automatically from text, techniques must be implemented for automatically reading texts into the computer. Currently, increasing numbers of texts, including equipment descriptions, software manuals, maintenance handbooks and even technical journals, are either composed by running punched tapes through type setting machines or printed by using computer-controlled CRT phototype setting techniques. In these cases the complete text is available and can be automatically read by the computer. This encourages the development of extraction of technical terms.

Both common and technical terms are read into the machine. Since only technical terms will be translated, it is necessary to eliminate the common vocabulary from the text. This can be done by storing a list of common words in the computer and comparing the text to this list. All the words for which a match is found are eliminated from the text. The remaining text thus consists solely of technical terms, which are then input into the retrieval procedure of the terminology bank to access their target equivalents.

We should note the fact that automatic text analysis also allows the isolation of technical terms which do not yet belong to the data base. Thus, such a program has two purposes: to produce terms to be translated and to find new terms to add to the data base. This decreases the time to locate new terms significantly.

Ligymom: Experimental on-line computer aids for the human translator (cf., Ling

Experimental computer aids for the human translator are being developed which basically consist of storage retrieval, editing and formatting operations carried out on line with a computer by an experienced human translator during the time in which a translation is produced. The system is not programmed to simulate the human translator by producing automatic translations. Rather, the user can call upon the computer's resources as needed in the translation process to shorten the delay between the initiation of a translation and production of a finished version. A combination of display terminals, computer hardware and software is used to perform functions which have habitual human counterparts of a mechanical nature, e.g., dictionary look-up, dictionary updating, creating of text-related glossaries, editing and layout, collection of text statistics, combination, insertion, and deletion of text. An essential aspect of this system of computer aids is that, while assuming the burden of much of the mechanical drudgery associated with production of a translation, it leaves to the translator those tasks whose successful completion is most heavily dependent on characteristics that are uniquely human, in particular, the ability to produce grammatical output in which appropriate target translations have been selected on the basis of understanding of text content rather than through heuristics of brute force.

The system is designed to make it maximally simple for work in an on-line environment by inexperienced computer users such as translators, terminologists, lexicographers, editors, and typists. The translation aids are implemented as modules which are compatible with existing text processing systems. As such they can either be integrated into such systems or isolated and put to other language processing uses with minimal modification.

The goal of the experimental computer-aided translation system is to streamline the entire translation process from the reception of a source text to the printing of the finished version of the translation, thereby significantly increasing the productivity of the translator. In this connection, the user can perform the following tasks on line:

1. Enter and/or edit a text, e.g., a translation or a dictionary.
2. Look up dictionary entries and browse through dictionaries and other reference files.
3. Update dictionaries or other text files.
4. Print text in formatted or unformatted layout.
5. Obtain text-related glossaries in textual word order or alphabetically sorted.
6. Obtain statistical information and concordances on translations and/or (machine-readable) source language texts.
7. Delete, merge, and duplicate text files or text portions.
8. Permit other users to share texts and dictionaries on-line and/or off-line.
9. Obtain instructions on how to use the system.

Expected advantages include:

1. Increased productivity through automated dictionary and terminology lookup, rapid and convenient revision of successive translation drafts, and high-speed layout and printing of translations;

2. Easily activated production of text-related glossaries, which can be saved for future work;

3. Maintenance of consistency in terminology through immediate accessibility of standardized terminological objects;

4. Easily activated automatic insertion of previously-translated text portions;

5. Reduced handling and consumption of paper through emphasis on the use of visual displays rather than printed output during all but the final processing phase.

An incremental approach to aids to translation (cf., Kay)

Existing tools already available on computers can be very useful for the day-to-day work of a translator. One is called text editing. Through a terminal a user can compose a text, delete some words, correct mistakes, etc.... The use of such tools is described in (Kay). At the same time the user can access the terminology bank to get information about a term, such as definitions or usage samples. The translator does his work on the terminal. After the text to translate is read into the computer, he displays a paragraph on the terminal, translates it, types in the translation, and then displays the next paragraph for translation. Here, interesting tools can be developed. For instance, a translator wants to translate a term consistently in a given text. Once the translator has made up his mind about the translation of a term, the computer system can then automatically point to the places, further in the text, where this term occurs. However, to ensure consistency in the updating of the translation, morphological rules are required to recognize the term when it occurs in different inflections, and these rules should exist in the source and the target languages. Algorithms for this are already in existence.

Kay advocates an incremental approach to machine-aided translation. The idea of the incremental approach is to add to the existing system only tools which are known to work well. This should eventually result in a system in which the computer performs the entire translation. At this point, the translator should still always be able to intervene in the translation process to the extent that he himself specifies. The computer should report any decision it makes, to avoid 'cascading errors.' Since the translator has control over the computer's decision, its bad guesses are avoided. This saves computer time and insures the quality of the translation. Martin Kay's ideas on translator intervention are an integral part of the design of the TARGET interactive system at CMU; at any point of the translation process the translator has access to the target text, to edit it into any desired form.

Myrtle Kay and Edward Lissman also contributed to the design of the TARGET prototype system being developed during 1977 at Carnegie-Mellon University.

The TARGET prototype system

Several points following the recommendations of the ELPAC report are being stressed in the development of this experimental system of ... [illegible]. Figure 5 lists the major components of a typical terminology bank while Figure 6 displays the components of the first, simpler form of the TARGET system.

A testable system

Consideration is being given to designing a testable system. System validation by formal experimentation may be used to discover what aspects of the system are useful, and to what types of translators. For example, the initial prototype TARGET system is validated by selecting several articles whose terminological content is placed in the bank. These articles are then translated by several professional translators, both with and without the system. A comparative evaluation is then made. Subsequent systems can be validated in a similar manner. An appropriate experimental design will factor out the various effects. Additionally, statistics can be kept on the frequency of use of the various aspects of the system.

A flexible approach to system design

A flexible approach to system design is implemented by the development of a system composed of separate units, or modules. Each module may operate independently or be incorporated into a larger routine. A standard terminology bank system can be planned first; this system alone might have lookup via terminal,

Interactive facilities

Then an interactive translation system can be introduced. The basic philosophy is to present the translator with a large set of facilities and allow him or her to use them as need fit. The development and refinement of facilities is continuing and ... [illegible] experience with the system is gained. The translator may use the computer to access information for selected terms, to generate a task-related glossary organized to his needs, or for interpersonal communication with other colleagues.

In the system, a translator may decide to create a term to be added to dictionary or update an old one. The creation and updating of the term via terminal in a direct access mode is a powerful tool for the terminologist-translator. He can directly review, supplement, modify or delete items in the data base. This, of course, raises protection problems. Not everybody should be allowed to change entries in the data base because their additions may be incorrect. Also, if several terminologists were able to file or alter terminals, steps must be taken to ensure that an entry is not attained by two people simultaneously. All modifications should be recorded on files which are filed and remain available for subsequent investigation of the history.

Figure 5. Main Components of a Terminology Bank

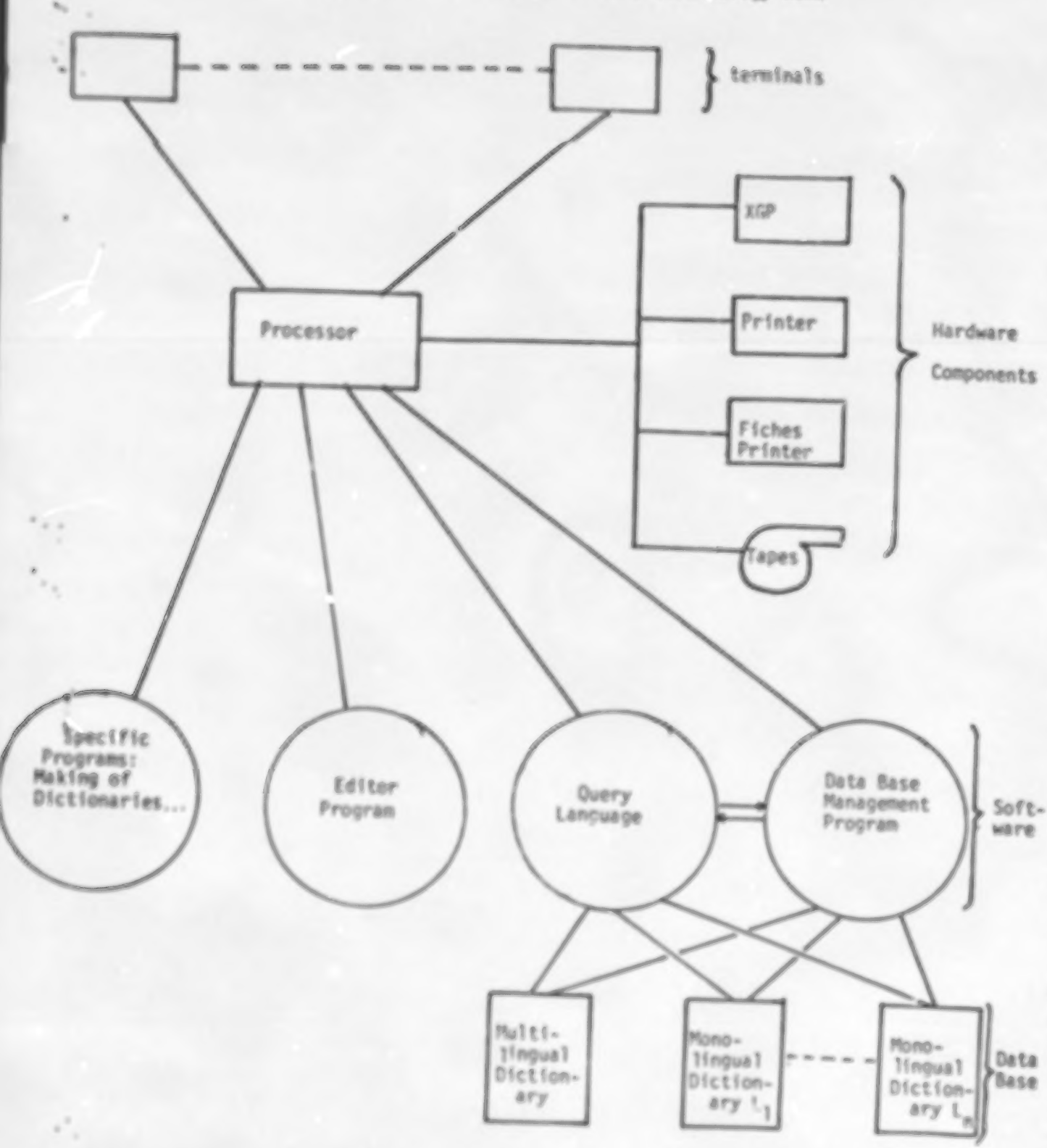
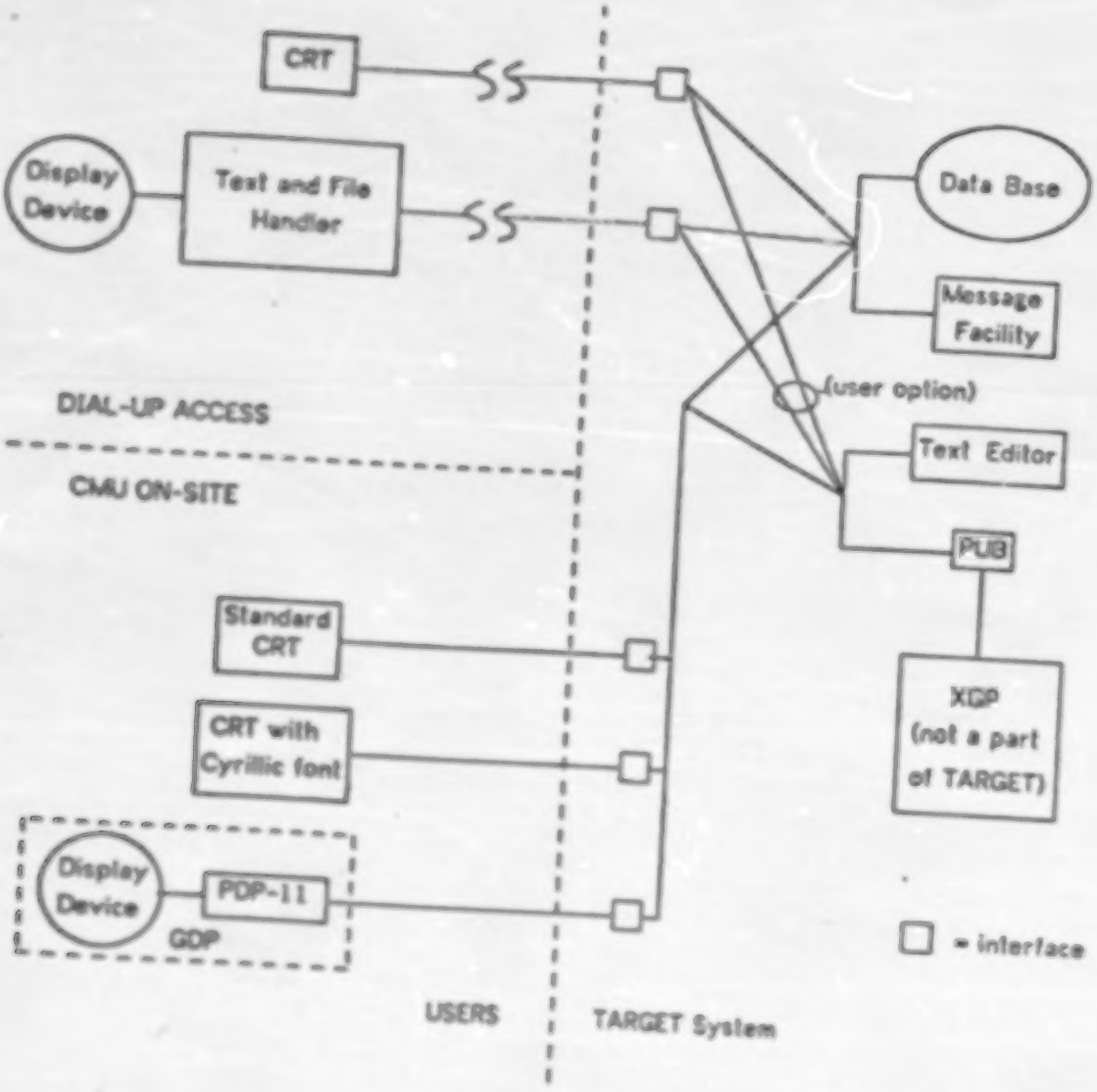


Figure 6. The Proposed TARGET Initial Configuration



The translator may decide to access dictionary definitions and type in the target text, editing it as he goes along. He may also type in the source text (or have it typed in), access dictionary definitions in the work space, and type the correct terms into the ongoing target translation, editing either the source or target representations as it seems desirable.

The following tasks are made possible by an especially designed editor. The translator may:

1. select among different target terms offered for an ambiguous source term,

2. edit awkward phrases,

3. ask for term definitions and synonyms,

4. ask for former translations of that term in context,

5. ask if the term occurs with above average frequency in this article (an indication of the topic of the article), and

6. specify any special uses of the term (for instance, that it be the standard term for correspondence on die casting with the Zaks Corporation).

For this purpose the translator interacts with the system through a terminal consisting of a keyboard, a screen, and censors to point at individual words and characters on the screen. The screen is divided into distinct regions called "windows," that is, definite areas serving several purposes.

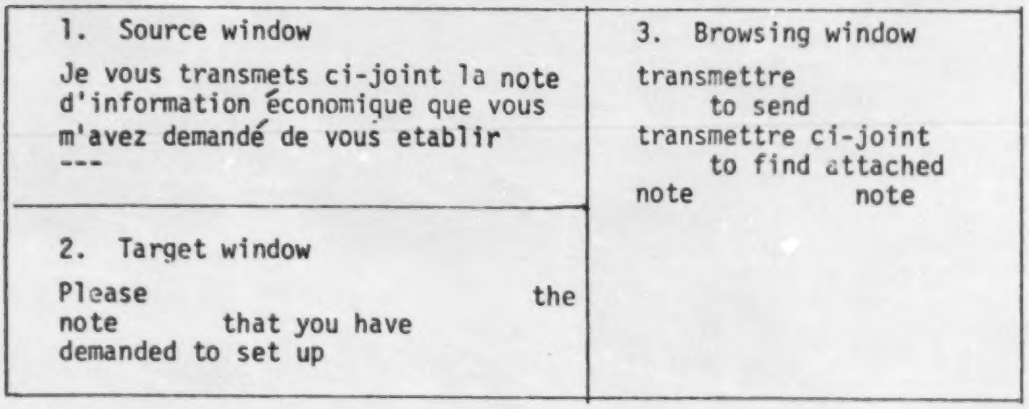
1. Source text window: the user accesses the source text through this window. It may be entered (and saved between sessions in a secure file system), or it may be originally obtained in machine-readable form from some external source. The user may make marginal notes and put "marks" in it. The source text may be perused by simply running through it by searches made for occurrences of specific terms.

2. Target text window: this is used to construct the translation. As with all other windows a font appropriate to the material displayed is used; the font may be changed by the user wherever desired, whether for emphasis (e.g., using an italic or bold font or one of a different size) or for linguistic purposes (where the orthography of a language is not expressible in a Roman font). Additionally, the user may perform many simple formatting operations, such as justification, underlining, and pagination.

An important feature of the source and target texts is that the translator may insert "marks" in them so that an annotation in one may be connected to some portion of another.

3. Browsing window: this window contains term-related information obtained from the terminology data base. Translators may wish to preserve the results of recent searches in this window. This window is used to display queries to the terminology data base; it is possible to call only for a term and its equivalent, or for any additional information (definition, source) associated to that entry. If the translator finds the right equivalent, he points to it with a cursor, causing the term or phrase to appear in the target text. In this window the results of recent searches can also be preserved.

Figure 7. Possible Screen Display during Translation Protocol 1



A long term goal is to create an editing option so that the translator types in the source document, and the computer automatically selects target equivalents and presents the provisional target text to the translator on a phrase-by-phrase basis for his approval. This is similar to an automatic translation routine in which post-editing has been moved into a more useful place; it is now at the source-target interface. Here, ambiguities and misinterpretations can be caught before they are compounded by later processing in the target language. However, we anticipate that this goal is beyond the phases described above.

Issues in Man-Computer Communication

A number of problems concerning man-computer communication must be faced before an aids-to-translation system may become useful to a user translator. None of the existing banks is used on-line by translators. The main reason is that the term banks' systems have been designed for specialists and are accessed only by specialists, that is by terminologists who act as intermediary links between the data bases and the translator. This state of things can be overcome if the user is provided with a computer environment which is not only adequate but also enjoyable to work in. Only if substantial care is given to designing systems with which the user can communicate can the difficulties now present in man-computer communication decrease, making computer technology really available to users.

The following are desirable characteristics of a language data base from the point of view of a common user:

1. Simplicity is the key characteristic for the creation, editing, and display of terms.

2. Flexibility is another desirable property, implemented in a system capable of incorporating modifications (e.g., a new language) and information from other systems. However, flexibility is secondary to simplicity.

3. Efficiency refers to the system's capability to offer search strategies with which the user is not acquainted.

Of the above criteria, simplicity is certainly the dominant one. It means that the system:

1. must be easily accessible by short clear rules. It is unnecessary and undesirable to require a user to learn a multiplicity of codes in order to communicate with a program. Similarly, the number of procedures for the user to learn must be reduced.

2. must offer as much support as needed by those unfamiliar with a computer. This means that commands should be accompanied by simple explanations. For example, if the command is "Regenerate," it should be followed by the explanation, "to correct dictionary order." While codes can be used to indicate fields of classification (ex. EC = Economics) for those familiar with the system, in the display of terms for the translators', names of fields should be spelled out.

3. must remain controllable by the user at all times; for instance, the user should be able to exit at any moment.

4. must seem natural for the non-technical user; particularly, the format of the display must match the user's expectations.

Particular care was given to the user's phase in the construction of the TARGET prototype system at CMU. Of the above criteria, TARGET stressed in particular 2 and 3. As many facilities as possible were made available to translators. They are described below:

> The user may employ abbreviations of commands, or even misspell them. In the case of ambiguity, s/he is presented with the ambiguity set so that the next attempt will be sure to be unambiguous. In addition to eliciting the set of alternatives at the current choice-point by typing "?," the user also has access to a number of help texts while remaining in the system. The user may configure some aspects of the system in accordance with his or her preference (e.g., the choice and sequence of term-fields presented upon the interactive creation of a new term). The user may augment choice sets where it is possible (e.g., the set of citations the program is to search). The system may store a "profile" if the user so desires, so that customization may occur at some suitable time (e.g., start-up) merely by reading in a command file. Such a profile, or set of profiles, would be part of a user's private workspace. [cf., GreE, p.6]

Between March and June 1977, translators were involved in testing this system. In these sessions it became clear that the criteria of offering a variety of supportive facilities to the user often conflicted with the criteria of naturalness. Commands like "well-done" or "do you want more?" often seemed time-wasting to the user. What follows is an indicative list of the modifications requested by the users; all seemingly related to the criteria of naturalness.

1. Simplification of language
   a. the word "note" is more natural than "gloss";
2. Simplification of supportive messages
   a. the "well done" message was objected to on the basis of time-waste;
   b. listing of facilities such as "cliche" which were of no immediate interest to the user were rejected;
   c. overly specialized "help" messages were rejected;
   d. messages "did you mean" and "may I suggest" were rejected because they applied only to a small fraction of errors;
   e. usage samples were not found useful.
3. Simplification of the display format
   a. clutter on the display was eliminated so that only the necessary information would appear displayed in specific parts of the screen.

For instance, the following display formats were preferred by translators during the design of the TARGET system. Three displays were incorporated into the translation program. In this program, initial information is typed in to access a term. The information includes the term, source, and target languages. If a term is accessed which belongs to more than one field, TARGET DISPLAY 1 is shown. The field is spelled out and placed on the right upper part of the screen.

TARGET DISPLAY 1

Header valve	Foreign trade Oil drilling

After the field is chosen, TARGET DISPLAY 2 is shown. This contains minimal information - only the equivalent.

TARGET DISPLAY 2

Header valve	Verbindungsventil

The translator glances quickly at it; if he wants more information, he types "more" and retrieves the rest of the information entered for the term. The result is TARGET DISPLAY 3.

TARGET DISPLAY 3

Header valve — Verbindungsventil

noun
"A valve with 3 heads." Websters New Coll.

CMU 1977

These DISPLAYS answer the criterion of simplicity and naturalness at the same time that they offer additional supportive facilities.

APPENDIX I. SAMPLES OF MECHANICAL TRANSLATION

FOUR DIFFERENT MECHANICAL TRANSLATIONS OF THE FIRST SENTENCE OF A RUSSIAN ARTICLE ON SPACE BIOLOGY [cf., Nat]

Bunker-Ramo Corporation No. 1

Biological experiments, conducted on various/different cosmic aircraft, astrophysical researches of the cosmic space and flights of Soviet and American astronauts with the sufficient/rather persuasiveness showed/indicated/pointed, that momentary/transitory/short orbital flights of lower/below than radiation belts/regions/flanges of earth/land/soil in the absence of the raised/increased/heightened sun/sunny/solar activity with respect to radiation are/appear/arrive/report safe/not dangerous/secure.

Computer Concepts, Inc. No. 2

The biological experiments that were carried out on different cosmic flying apparatus, ASTROFIZICESKIE the research of cosmic PROSTRANSTVA and the flights of Soviet and American KOSMONAVTOV with sufficient UBEDITEL6NOST6H showed, that the short-time orbital flights below of the radiational belts of earth in the absence that was raised by the SOLNECNOI one of the activity in a radiational attitude are BEZOPASNYMI.

FTD, USAF No. 3

Biological experiments, conducted on different space aircraft/vehicles, astrophysical space research and flights of Soviet and American astronauts with/from sufficient convicingness showed that short-term orbital flights lower than radiation belts of earth in the absence of heightened solar activity in radiation ratio are safe.

EURATOM, Ispra, Italy No. 4
(Essentially the Georgetown MT system)

Biological experiments, which were conducted on different cosmic LETATEL6NYX APPARATI, the astrophysical investigations of cosmic space and the flights of Soviet and also American KOSMONAVTOV with the sufficient convincingness showed, that the short-term orbital flights of below radiation belts of ground upon the absence of the increased solar activity in radiation relation are safe.

REFERENCES

[Akh] Akhmanova, Olga, and Agapova, Galina. Terminology: Theory and Method. Moscow: Moscow State University, 1974.

And Anderla, J.G. Information in 1985: A Forcasting Study of Information Needs and Resources. Organisation for Economic Cooperation and Development, Paris, France, 1973. p. 123.

Arc Archimbaud, Robert. "Presentation et Protocole d'Utilisation d'un Nouveau Système de Classement des Données Terminologiques et Documentaires." Ottawa: Secrétariat d'Etat, Bureau des Traductions, December 1976.

Bac1 Bachrach, J.A. "L'Ordinateur au Service du Terminologue. Maître ou Esclave [The Computer for the Terminologist's service. Master or Slave.]" META, Journal des traducteurs / Translator's Journal, 16 (1971): 105-114.

Bac2 Bachrach, J.A. "An Experiment in Automatic Dictionary Look-up." The Incorporated Linguist 13, no. 2 (April 1974): 47-49.

BaN Baumann, E., and Neubert, G. "Das Elektronische Worterbuch der Fachsprachen - Aufgaben und lexikographische Struktur [The computerized vocabulary of specialized languages - Tasks and lexicographic structure.]" Wissenschaftliche Zeitschrift der Technischen Universität Dresden 23, no. 3/4 (1974): 627-633.

Bel1 Belahov, L.Ju. Standartizacija Terminologii v SSSR i za Rubežom. Obzor [Standardization of Terminology in the USSR and Abroad. A Synopsis.] Moskva: Izd-vo Komiteta standartov, mer i ixmeritel' nyh priborov pri Sovete Ministrov SSSR, 1966.

Bel2 Belahov, L.Jü.; Popov-Cerkasov, I.N. Problemy Gosudarstvennoj Standartizacii Terminologii v SSSR [Problems of Standardization of Terminology by the State in the USSR] Moskva: Izd. Komiteta Standartov Mer i Izmeritel'nyh Priborov pri Sovete Ministrov SSSR, 1968.

Bri1 Brinkmann, K.H. Uberlegungen zum Aufbau und Betrieb von Terminologie - Datenbanken als Voraussetzung der maschinenunterstützten Übersetzung. [Considerations on the Construction and Operation of Terminological Word Banks as Prerequisite of Computer Aided Translation] . Nachrichten fur Dokumentation 25 (1974), no. 3, p. 99-104.

Bri2 Brinkmann, K.H., "The TEAM Program System." Philips Terminology Bulletin 4, no. 2/3 (June 1975).

BriT Brinkmann, K.H., Tanke, E. "The TEAM Program System and International Cooperation in Terminology." In: International Co-operation in Terminology. Infoterm Series 3. First Infoterm Symposium. Munchen: Verlag Dokumentation, 1976.

Can1 Canada. Québec. l'Office de la Langue Française. Guide de Travail en Terminologie. Cahiers de l'Office de la Langue Française, no. 20. Québec, 1973.

Can2 Canada. Québec. l'Office de la Langue Française. "La Normalisation Linguistique." Actes du Colloque International de Terminologie. Québec: 1974.

Can3 Canada. Secrétariat d'Etat. Bureau des Traductions." Direction generale de la Terminologie et de la Documentation." Colloque Canadien sur la Fondements d'une Methodologie Generale de la Recherche et de la Normalisation et Terminologie et en Documentation Premier Partie: Text des Exposés. Ottawa: Feb. 16-18 1976.

Can4 Canada. Terminology and Documentation Directorate. Standardization and Laison Branch. The Terminology Bank of Canada; an Overview. Ottawa: January 1977.

Cha Chapin, Paul G., "Report on Seminar on Machine Translation." In Federal Broadcast Information Service. "FBIS Seminar Proceedings." Rosslyn,Virginia, March 8-9, 1976. Copies can be obtained from J. Mathias, 9811 Connecticut Ave., Kensington, Md. 20795.

Con Conseil International de la Langue Francais (CILF). La Clé des mots. Cahiers de Terminologie. Paris: CILF. A monthly publication.

Com Commission des Communantés Européenes. Translation Division. Terminology Bureau. System of Classification, compiled by H. Lenoch. Doc. 3642/73. Luxembourg, 1975.

DNA DNA. Der kurze Weg zur Datenbank. [The Short Way to the Data Bank] DIN Informationssystem Technik (DINST). Berlin: DNA.

Dub Dubuc R., "Formation des terminologues. Theoriciens ou praticiens." Colloque ATIO, Ottawa, Sept. 1974.

Fao FAO. "Inter-Agency Arrangements for Terminology Exchange and Improvement, Including the Use of Computer Technology." Rome: Sept. 1976.

Fel1 Felber, Helmut. "Outline of a Study on the Development of a Network of Terminology Preparing Bodies and Terminology Documentation Centres." fe/vt 1976 01 03. Vienna: International Information Center for Terminology. (Infoterm), 1976.

Fel2 Felber, Helmut. "Standardization of Terminology in USSR. A cursory review." Babel 16, no. 4 (1970): 212-214.

Fel3 Felber, Helmut. "Developing International Co-operation in Terminology and Terminological Lexicography." Vienna: International Information Center for Terminology (Infoterm).

Fel4 Felber, Helmut. "An Outline of International Terminological Activities." Vienna: International Information Center for Terminology (Infoterm).

Fel5 Felber, Helmut. "International Standardization of Terminology. Theoretical and Methodological Aspects." Vienna: International Information Center for Terminology (Infoterm).

Fed Federation Internationale des Traducteurs. "Project de Resolution en Vue d'Etablir une Fiche d'Echange Terminologique Recommandée par la Federation Internationale des Traducteurs." VII Congrès de la FIT. May 1974.

For Fortin, J.M. "La Banque de Terminologie du Québec." Québec: Banque de Terminologie. May 10, 1974.

Fra Francois, Pol. "La Banque de Données Terminologiques EURODICAUTOM de la Commission des Communautés Européennes: Le Resultat d'un Travail Pluridisciplinaire." Presented at Colloque International de Terminologie, Paris - La Defense, June 1976. Luxembourg: Commission des Communautés Européennes. Translation Division. Terminology Bureau.

FrP Fredericksen, B., and Power, L. "A Query System for Reviewing On-line Manuals." Computing Center Newsletter. IBM T.J. Watson Research Center, Yorktown Heights, New York. Vol. 8, no. 4 (March 10, 1975): 27-29.

Ger Gerrard, Martha W., "Regular Use of Machine Translation of Russian at Oak Ridge National Laboratory." In Federal Broadcast Information Service. "FBIS Seminar Proceedings." Rosslyn, Virginia, March 8-9, 1976. Copies can be obtained from J. Mathias, 9811 Connecticut Ave., Kensington, Md. 20795.

Goe Goetschalckx, J. "Die Terminologiearbeit der Sparchendienste der Europäischen Gemeinschaften. [Terminology Work Carried Out by Language Services of the European Communites]." In: Fachseminar Terminologie und Lexikographie des Landesverbandes Hessen, Oct. 23-25, 1974. Bundesverband der Dolmetscher und Übersetzer. Frankfurt/Main.

Gos Gosstandart SSSR. VNIIKI. Avtomatizirovannaja Sistema Informacionnoterminologicheskogo Obsluzivanija. [Automated System of the Terminological Information Service]. Moscow: Gosstandart SSSR, 1974.

GreC Greenfield, Concetta C.; Erman, Lee; Burge, John; and Morton, Sara. "Carnegie-Mellon University's Terminological Data Bank Project TARGET: Its History, Goals and Current Status." Translation Center and Department of Computer Science, Carnegie-Mellon University, Feb. 1977.

GreM Greenfield, Concetta C., and Morton, Sara. "Research Proposal for a Machine-Aided Translation System." Translation Center, Carnegie-Mellon University, July 1976.

GreS Greenfield, Concetta C., and Seroto Daniel. "Report of Visit to European Terminology Banks." Translation Center, Carnegie-Mellon University, Jan. 1977.

GreT Greenfield, Concetta C., et. al. "Target Inter-office Correspondence." Translation Center, Carnegie-Mellon University, July 1977.

Hor Horn, J. Computer and Data Processing Dictionary and Guide. Englewood Cliffs, New Jersey: Prentice-Hall, 1966.

Int International Atomic Energy Agency. "Information Systems: Their Interconnection and Compatibility." Vienna: 1975

Iri IRIA: Hyafil; Rohmer; Serain, Tuseve, Van de Wiele. "Rapport d'Etude sur la Machine de Consultation de Données." (To be published). IRIA/France: Laborie

Iso1 ISO (International Standards Organization). "Vocabulary of Terminology." ISO/R 1087. Geneva: 1969

Iso2 ISO (International Standards Organization). "Les Relations [associations] Entre les Notions [concepts]." ISO/TC 37 (Secretariat - 151) 251, 1974-06. Geneva: 1974.

Iso3 ISO (International Standards Organization). "Symboles Lexicographiques Particulièrement pour l'Emploi dans les Vocabulaires Systematiques à Definitions." Norme Internationale ISO 1951. Ref. no ISO 1951-1973 (F) Geneva: 1951.

Iso4 ISO (International Standards Organization). "Terminology (Principles and Coordination)". Fifth meeting of ISO/TC 37, Vienna June 1974. Geneva: 1974.

Iso5 ISO (International Standards Organization). "Format for Terminological/ Lexicographical Data Interchange on Magnetic Tape." ISO/TC 37/WG 4 N 19 E.

Jor Jordan, Sara. "Report on FBIS Seminar on Computer Aids to Translation." In Federal Broadcast Information Service, "FBIS Seminar Proceedings," Rosslyn, Virginia, March 8-9, 1976. Copies can be obtained from J. Mathias, 9811 Connecticut Ave., Kensington, MD. 20795.

Kan Kandelaki, T. L. "Značenija Terminov i Sistemy Značenij Naučno-tehničeskih Terminologij [The Meaning of Term and Systems of Meanings of Scientific and Engineering Terminologies]." In: Problemy Jazyka Nauki i Tehniki (1970): 3-39

Kay Kay, Martin. "The Proper Place of Men and Machines in Language Translation." Amer. J. Computational Linguistics, AJCL Microfiche 46, 1976.

Krol Krollmann, F. "Linguistic Data Banks and the Technical Translator." Meta 16, no. 1/2 (1971): 117-124

Kro2 Krollmann, F.; Schuck, H.; and Winkler, U. "Herstellung textbezogener Fachwortlisten mit einem Digitalrechner - ein Verfahren der automatischen Übersetzungshilfe." Beitrage zur Sprachkunde und Informationsverarbeitung 6 (January 1966) 7-30.

Kul Kulebakin, V. S. Sovremennye Problemy Terminologii v Nauke i Tehnike. [Present Problems of Terminology in Science and Engineering]. Moscow: Izd. Nauka, 1969.

Kuf2 Kufaev, V.S. Kak fotokat'not terminologiei. Domoy i Metody [Terminological Work: Theory and Methods]. Moskva: Izd-vo Nauka, 1968. p. 63-75. A bibliographical list of all publications prepared by the Committee on Scientific and Technical Terminology of the Academy of Sciences of the USSR.

Lan Lang, F. H. Automatisierte Herstellung von Thesauren und Begriffssystemen für Wörterbücher und Informationssprachen [Automated Establishment of Thesauri and Systems of Concepts for Vocabularies and Specialized Terminologies]. Nachrichten für Dokumentation 24, no. 6 (1973): 232-236.

Lau1 Laurent, J.S.F. "La Banque Automatisée de Données Terminologiques de l'Office NORMATOM." Courrier de la normalisation (Paris), 42, no. 248, (1973): 402-406.

Lau2 Laurent, J.S.F. "NORMATOM: Sa Conception, son Exploitation et sa Place dans Réseau International d'Information sur les Normes." In: Deltaform Series 3, Premier Symposium d'Deltaform. Coopération Internationale en Terminologie. Munchen: Verlag Dokumentation, 1976, p. 193-223.

Laj Lejčik, V.M.; Selinov, I.P. "Oblasti Peresečenija Terminovedenija i Perevoda Naučnoj i Tehničeskoj Literatury [Fields of Overlapping between the Theory of Terminology and Translation of Scientific and Engineering Literature]." Naučno-tehničeskaja informacija, Serija 2, no. 12 (1971): 20-25.

Lem Lemcke, H. "Klassifikationssysteme [Classification Systems]." In Terminologie Bulletin 28. Luxembourg: Commission des Communautés Européennes,Bureau de terminologie, p. 93-134.

Lip Lippmann, Erhard O., "On-line Generation of Terminological Digests in Language Translation: an Aid in Terminology Processing." IBM Research communication, Sept. 1975.

Mac Machine-Assisted Translation in West Germany. JPRS 68710. Springfield, Virginia: U.S. Joint Publications Research Service, March 1977.

Mar Martin, J. Computer Data-Base Organization. Englewood Cliffs, New Jersey: Prentice-Hall.

Mat Matthias, J., and Fender, M. "Machine-Aided Translation Experiment." (July 20, 1972). Copies can be obtained from J. Matthias, 3811 Connecticut Avenue, Washington, DC. 20008.

Mcd McDavid,Raven I. Jr., and Duckert, Audrey R., eds. Lectionography in English. Annals of the New York Academy of Sciences, vol. 211. New York: New York Academy of Sciences, 1973.

Mon1 Montréal, Université de Banque de Terminologie. Description du Système TERMIUM. By Roberto Dubuc. Jan. 1975.

Mon2 Montréal, Université de Banque de Terminologie. Répertoire des Domaines avec Leurs Codes Respectifs / Fields Classification and their Individual Codes. Montréal: 1976.

Burton, Dana. "Lexicographic Considerations for Term Entry: How Should Term Meaning be Conveyed?" Translation Center, Carnegie-Mellon University, Aug. 1976.

Burton, Dana, and Burge, John. "Term Connectives in a Data Base." Translation Center, Carnegie-Mellon University, March 1977.

Kocourek [illegible] Deutschsprachige Universität, Fifteenth-Kocourek Fakultät [illegible] Reading alternative "Semantisch-syntaktische Probleme [illegible] Informatik." Scientific Symposium "Semantic Problems of the Language of Science, Terminology and Information" [in 2 parts] 2 Vorträge. Review: Internat' cour Modern-Usage Universitäts, 1977.

National Academy of Sciences, National Research Council, Automatic Language Processing Advisory Committee. Language and Machines: Computers in Translation and Linguistics. Publication 1416. National Academy of Sciences, National Research Council, Washington, D.C., 1966, p. 34.

Neubert, S., et al. Das arbeitswissische Wörterbuch der Technischen - mechanotechnische Terminologie. Linguistic and Subject Time Connection and vocabulary of subject-area languages - computational processing. Input and output. Proc. 2. der I.A. Dresden 23, no. 3/4 (1976).

Paré, Marcel. "Comment les Banques Automatisées de Terminologie Multilingues Peuvent Accroître Les Ergonomies de Traduction." In Colloque Sur la Coopération Internationale en Terminologie. Paris: 1976.

Paré, Marcel. "Les Banques Automatisées de Terminologie Multilingues et les Ergonomies de Traduction." In Infoterm Series 3, Premier Symposium d'Infoterm. Coopération Internationale en Terminologie. München: Verlag Dokumentation, 1978.

Paré, Marcel. "The University of Montreal's Banque de Terminologie." International Federation of Translators. Bled: May 1976.

Pozzo-Cardasso, I.R., and Veltcheva, I.R. "Mashinnovedsko [illegible] Rechnik i Terezi: [illegible] Terminological Documentation Systems in [illegible] Film of the Language of Engineering Documents Prepared by the Members of the Council for Mutual Economic Aid." Machine-[illegible] Terminologie. Review 10, no. 4 (1977): 36-45.

Reichling A. "Terminological Data Bank of the Commission of the European Communities : Luxembourg." In Proceedings of the International Colloquium on Terminology. Paris: La Defense, June 1976.

Regie de la Langue Française. Banque de Terminologie. Thesaurus de Domaines d'Usage. Québec: 1975.

Rieger C.J. "Overview to the FEDS Seminar." In Federal Research and Information Service. "FEDS Seminar Proceedings." Reston, Virginia, March 8-9, 1976. (Copies can be obtained from J. Rubinow, 8601 Connecticut Ave., Kensington, MD. 21795.

Roh Rohaert, A. "Vers un Reseau International de Terminologie." In: Infoterm Series 3. Premier Symposium d'Infoterm. Cooperation International en Terminologie. Munchen: Verlag Dokumentation, 1976.

Sag Sager, J.C. Classification and Hierarchy in Technical Terminologies.

Sch Schulz J. "A Terminology Data Bank for Translators. Methods of Interrogation in the TEAM system." Munich: Language Services Department, Siemens AG.

Ser Serain, Daniel. "Design of the Data Base for the Terminology Bank." Translation center, Carnegie-Mellon University, Sept. 1976

Sjo Sjöstrand, N.G. "Tekniskt Nomenklaturarbete i Sverige [Technical Terminology Work in Sweden]." TNC Aktuellt (Stockholm) 11, no. 5 (1969): 3-4.

Spe Speers, Peter C. Oil Industry Technical Terms in Arabic. First Arab Petroleum Congress. Cairo: April 1959.

Sun Sundström, E. "Word Bank Building and Terminological Policy." In Infoterm Series 3 First Infoterm Symposium. International Co-operation in Terminology. Munchen: Verlag Dokumentation, 1976.

Tan1 Tanke, E. "Electronic Data Processing in the Service of Translators, Terminologists, and Lexicographers." Philips Terminology Bulletin 4, no. 2/3 (June 1975).

Tan2 Tanke, E. "Future Developments [for Siemens AG]." Philips Terminology Bulletin 4 no. 2/3 (June 1975).

Uni UNISIST Steering Committee. "Report on the Meeting for the Evaluation of the International Information Centre for Terminology." Paris: August 1976.

Vni1 Vniiki. "Informacionno-Poiskovaja Sistema na Mašinnom Slovare Standartizovannoj i Rekomendumoj Naučno-Tehničeskoj Terminolgii [Computer Retrieval System for Vocabulary of Standardized and Recommended Scientific and Technical Terminology]." Naučno-tehničeskaja Terminologija 1 no. 45 (1974): 41-43.

Vni2 Vniiki. "Spravočnyj Bank Terminov Avtomatizirovannoj Sistemy Informacionno-Terminologičeskogo Obsluživanija (SBT ASITA) [Terminology data bank of the computerized system of the terminology information service (SBT ASITA)]." Naučno-tehničeskaja Terminologija 8 no.65 (1975): 3-12.

Voj Vojtko Margaret M. "Bibliography of French-English, English-French and Polyglot Dictionaries." Translation Center, Carnegie-Mellon University, July 1975.

Wil Wilks, Yorick. "A Small Machine Translation System Based on Deep Semantic Structures." In Federal Broadcasting Service. "FBIS Seminar Proceedings," Rosslyn, Virginia, March 8-9, 1976. Copies can be obtained from J. Mathias, 9811 Connecticut Ave., Kensington, Md. 20795.

Wils — Wilss W. The Science of Translation: An Analytical Bibliography. Vol. I, Tubingen: TBL, 1970; Vol. II, Tubingen: TBL, 1972.

Win — Winograd, Terry. Understanding Natural Language. New York: Academic Press, 1972.

Wus1 — Wuster, Eugen. "Die Allgemeine Terminologielehre - ein Grenzgebiet zwischen Sprachwissenschaft, Logik, Ontologie, Informatik und den Sprachwissenschaften. [The General Theory of Terminology - a Border Field between Linguistics, Logic, Ontology, Information Science and the Subject Fields] ." Linguistics (The Hague) 119 (1974): 61-106.

Wus2 — Wuster, Eugen. "Begriffs - und Themaklassifikation. Unterschiede in ihrem Wesen und ihrer Anwendung. [Concept and Subject Classifications. Difference between their Nature and their Application.]" Nachrichten für Klassifikation 22, no. 3 (1971): 98-104 and no. 4: 143-150.

Wus3 — Wuster, Eugen. Bibliography of Monolingual Scientific and Technical Glossaries Volume I. National Standards. Paris: Unesco, 1955.

Wus4 — Wuster, Eugen. Internationale Sprachnormung in der Technik, besonders in der Elektrotechnik [International Standardization of Languages in Engineering, Particularly in Electrical Engineering] Bonn: H. Bouvier & Co. Verlag, 1970.

Wus5 — Wuster, Eugen. The Machine Tool/Machine-outil. Vol.1. London: The Technical Press Ltd., 1968. Grundbegriffe bei Werkzeugmaschinen. [Basic concepts of machine tool] Vol. 2. (Grundbegriffe bei Werkzeugmaschinen. [Basic Concepts of the Machine Tool]). London: The Technical Press Ltd., 1968.

Wus6 — Wuster, Eugen. The Road to Infoterm. Two reports prepared on behalf of Unesco: Inventory of Sources of scientific and technical terminology; a plan for establishing an international information centre (clearing house) for terminology. Infoterm Series 1. Munchen: Verlag Dokumentation, 1974.

Zgu — Zgusta, Ladislav. Manual of Lexicography. Paris: Mouton, 1971.